

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO–MATEMATIČKI FAKULTET
MATEMATIČKI ODSJEK

Marija Višić

ODREDNICE PRORAČUNSKE
TRANSPARENTNOSTI GRADOVA I
OPĆINA U ZADARSKOJ ŽUPANIJI

Diplomski rad

Voditelj rada:
dr. sc. Katarina Ott

Zagreb, rujan 2017.

Ovaj diplomski rad obranjen je dana _____ pred ispitnim povjerenstvom u sastavu:

1. _____, predsjednik
2. _____, član
3. _____, član

Povjerenstvo je rad ocijenilo ocjenom _____.

Potpisi članova povjerenstva:

1. _____
2. _____
3. _____

Sadržaj

Sadržaj	iii
Uvod	2
1 Proračunska transparentnost: definicija, pregled literature i hipoteze	3
1.1 Pojam proračunske transparentnosti	3
1.2 Pregled literature	4
2 Teritorijalni ustroj i proračunska transparentnost u Republici Hrvatskoj	10
2.1 Sustav lokalne i područne samouprave u Republici Hrvatskoj	10
2.2 Proračunska transparentnost u Republici Hrvatskoj	11
3 Metodologija	14
3.1 Generalizirani linearni modeli	14
3.2 Prilagodba modela	18
3.3 Poissonov regresijski model	26
4 Podatci	33
4.1 Zavisna varijabla	33
4.2 Nezavisne varijable	33
5 Analiza i rezultati	37
5.1 Deskriptivna statistika	37
5.2 Korelacija	39
5.3 Jednostruka Poissonova regresija	40
5.4 Višestruka Poissonova regresija	41
5.5 Usporedba modela	55
5.6 Interpretacija rezultata	57
6 Zaključak	59

SADRŽAJ

iv

Bibliografija

61

Dodatak

67

Uvod

Proračunska transparentnost najčešće se definira kao "sveobuhvatno, razumljivo, pouzdano, pravovremeno i primjenjivo javno izvještavanje o prošlom, sadašnjem i budućem stanju javnih financija" (vidi [14]). Međunarodni je monetarni fond (dalje u tekstu: MMF) još 1998. razvio *Kodeks dobre prakse o pitanjima fiskalne transparentnosti* kao međunarodni standard za otkrivanje podataka o javnim financijama ([26]).

U Republici Hrvatskoj (dalje u tekstu: RH) Institut za javne financije (dalje u tekstu: IJF) proračunsku transparentnost mjeri brojem objavljenih ključnih proračunskih dokumenata na internetskim stranicama lokalnih jedinica. Taj je indeks nazvan OLBI (engl. Open Local Budget Index). Dosadašnja mjerenja pokazala su prilično nisku razinu OLBI-ja, ali se sa svakim ciklusom ta razina povećava (vidi [27]).

Stoga je prirodno da se pokušava otkriti uzrok niske razine proračunske transparentnosti, zašto su neke lokalne jedinice transparentnije od drugih i kako bi se razina proračunske transparentnosti mogla poboljšati.

Cilj je ovog rada otkriti odrednice proračunske transparentnosti gradova i općina u Zadarskoj županiji, i tako barem djelomično odgovoriti na pitanja postavljena u gornjem odlomku. Promatra se OLBI za razdoblje studeni 2015. - ožujak 2016.

U skladu s postojećom literaturom, promatraju se socio-kulturološke, ekonomske i političke varijable. Postavljene su sljedeće hipoteze:

- veličina lokalne jedinice, postotak starijeg stanovništva, postotak računalno pismenih, postotak kućanstava koja koriste Internet, razina obrazovanja, postotak muškaraca, politička konkurencija, izlaznost na izbore, odmak od proračunske ravnoteže, dohodak po stanovniku, proračunski prihodi po stanovniku i rashodi lokalne jedinice na računalne usluge imaju pozitivan utjecaj na OLBI
- reizbor (grado)načelnika negativno utječe na OLBI

Hipoteze se empirijski testiraju na uzorku od svih 34 lokalnih jedinica u Zadarskoj županiji - 6 grada i 28 općina. Koristi se regresijska analiza, konkretnije Poissonov log-linearni model unutar okvira generaliziranih linearnih modela jer je varijabla od interesa OLBI prebrojiva.

Između nekoliko podjednako dobrih modela, za najbolji je odabran model sa šest prediktora. Hipoteze su time djelomično potvrđene. Rezultati pokazuju da postotak kućanstava s Internetom, politička konkurentnost, odmak od proračunske ravnoteže i dohodak po stanovniku zaista imaju pozitivan utjecaj na OLBI. Za veličinu lokalne jedinice i izlaznost na izbore dobiveni su iznenađujući rezultati - u Zadarskoj je županiji utjecaj tih varijabli negativan. Ostale se varijable nisu pokazale značajnima u objašnjenju razine OLBI-ja.

Rad je strukturiran u veća poglavlja. Prvo poglavlje uključuje motivaciju za nastanak rada - objašnjava pojam proračunske transparentnosti, donosi pregled literature i razvija skup hipoteza temeljenih na teoretskim obrazloženjima i empirijskim rezultatima. U drugom poglavlju čitatelja se upoznaje sa sustavom lokalne i područne samouprave u RH, zakonskim odredbama, načinom mjerenja i stanjem proračunske transparentnosti u RH s naglaskom na lokalne jedinice Zadarske županije. U trećem se poglavlju predstavlja teoretski okvir unutar kojeg se u petom poglavlju provodi analiza. Četvrto poglavlje donosi detaljan opis varijabli korištenih u analizi. Zadnje poglavlje ističe najvažnije zaključke rada te donosi njegova ograničenja i preporuke za daljnja istraživanja i djelovanje vladajućih.

Poglavlje 1

Proračunska transparentnost: definicija, pregled literature i hipoteze

Kako bi se opravdalo provedeno istraživanje, nužno je razumjeti što je to proračunska transparentnost i zašto je ona važna. Prvi odjeljak ovog poglavlja objašnjava pojam proračunske transparentnosti, dok se u drugom odjeljku donosi širi pregled literature i konačno postavljaju hipoteze koje će se testirati u glavnom dijelu rada.

1.1 Pojam proračunske transparentnosti

MMF proračunsku transparentnost definira kao "sveobuhvatno, razumljivo, pouzdano, pravovremeno i primjenjivo javno izvještavanje o prošlom, sadašnjem i budućem stanju javnih financija" (vidi [14]). Proračunska transparentnost osigurava da vlade imaju točnu sliku o svojim financijama prilikom donošenja poslovnih odluka. Osim toga, ona zakonodavnom tijelu, tržištima i građanima osigurava informacije koje su im potrebne da bi vladu smatrali pouzdanom.

Budući da je proračunska transparentnost ključna za učinkovito fiskalno upravljanje i odgovornost, MMF je još 1998. godine razvio *Kodeks dobre prakse o pitanjima fiskalne transparentnosti* (engl. *Code of Good Practices on Fiscal Transparency*) kao međunarodni standard za otkrivanje podataka o javnim financijama. Skup načela i praksi razvijen je oko četiri stupa: jasnoća uloga i odgovornosti, otvoren proračunski proces, javno dostupne informacije i osiguranje integriteta fiskalnih podataka (vidi [26] i [4]).

Alesina i Perotti (vidi [1], 403.) smatraju da političari ne žele usvojiti najtransparentnije prakse. Kao razlog tome navode dva teorijska argumenta. Prvi, teorija fiskalne iluzije sugerira da se troškovi vlade čine manjima nego što zaista jesu kada proračuni nisu dovoljno transparentni ili kada ih porezni obveznici ne shvaćaju u cijelosti. Drugi, taktičkom

neodređenošću u otkrivanju proračuna političari stvaraju prednost za ostvarivanje svojih ciljeva čak i s racionalnim biračima. Dalje zaključuju da nedostatak transparentnosti, koji se očituje skrivanjem poreznih opterećenja, isticanjem prednosti potrošnje i podcjenjivanjem trenutnih i budućih financijskih obveza, dovodi do zbunjenosti o pravom stanju javnih financija.

1.2 Pregled literature

Kroz posljednjih nekoliko desetljeća u svijetu se velika pozornost počinje posvećivati proračunskoj transparentnosti. Radi se o prilično apstraktnom pojmu za koji je teško naći mjeru koja bi ga savršeno predstavljala, stoga postoji mnogo literature na tu temu. Istraživanja na lokalnoj razini ipak su rjeđa od onih na nacionalnom nivou, a u RH su takva istraživanja tek u začetima.

Indeks proračunske transparentnosti katalonskih lokalnih jedinica razvijen 2012. spada među najjednostavnije. Esteller-Moré i Polo Otero (vidi [11], 1158. - 1161.) zadovoljavaju se tek razlikovanjem transparentnijih jedinica od onih manje transparentnih po kriteriju dostavljanja zahtijevanih godišnjih izvješća do utvrđenog roka. Dostavi li lokalna jedinica više od pola traženih dokumenata, smatra se transparentnom, u suprotnome, ne.

Transparency International ide korak dalje pa za lokalne jedinice u Španjolskoj stvara dosta opsežniji indeks cjelokupne transparentnosti. Taj indeks sadrži podatke od prethodne tri godine i temelji se na prikupljanju podataka i analizi upitnika poslanih lokalnim jedinicama s pitanjima vezanima za pet različitih područja transparentnosti: transparentnosti poduzeća, društvene transparentnosti u odnosima s građanima, proračunske transparentnosti, urbanističkog planiranja i transparentnosti nabave te transparentnosti ugovaranja usluga (vidi [35], 95.).

Već je spomenuto da je teško naći prikladnu mjeru za proračunsku transparentnost, a samim time je koncept potražnje za njom još kompliciranije mjeriti. Stoga je indeks potražnje građana za transparentnošću, koji su konstruirali Piotrowski i Van Ryzin (vidi [28]), vrijedan spomena kao jedini te vrste. Podatci su prikupljeni *online* istraživanjem na dobrovoljnoj bazi pa ne predstavljaju reprezentativni uzorak ispitanika. Kako bi se smanjila pristranost, analiziraju se podatci ponderirani s obzirom na znane karakteristike populacije. Anketa ispituje mišljenje građana o tome kakve bi izvještaje i informacije lokalne jedinice trebale objavljivati te općenitije stavove vezane za odnose građana i vladajućih. Kako bi se brojna pitanja iz ankete sažela i grupirala, provodi se eksplanatorna faktorska analiza. Pokazalo se da se potražnja javnosti za transparentnošću očituje u više dimenzija - kroz

proračunsku transparentnost, dostupnost informacija o kriminalu i sigurnosti, načelo otvorene vlade, te transparentnosti i dobre i iskrene vlasti (vidi [28], 311.-315.). Analiza daje i odrednice tih dimenzija te stoga pruža dobre temelje za daljnje radove na tu nedovoljno istraženu temu.

Za ovo su istraživanje važniji načini mjerenja koji uključuju izvještavanje na Internetu. Internet je vjerojatno jedan od najpristupačnijih i najjeftinijih načina za širenje informacija u današnjem društvu, stoga se mjerenje sve više usredotočava na e-demokraciju i ulogu novih tehnologija. Kod analize izvještavanja te vrste također postoje indeksi različite složenosti.

Styles i Tennyson (vidi [36]) mjere dostupnost i pristupačnost financijskih izvještaja na internetskim stranicama lokalnih jedinica u SAD-u. Dostupnost je binarna varijabla, a pristupačnost je mjerena indeksom čije komponente su temeljene na preporučenim praksama za izvještavanje putem Interneta.

Grupa španjolskih ekonomista (vidi [6]) kreira tri indeksa: sadržaj pruženih informacija, kvalitativne značajke objavljenih informacija te navigaciju, dizajn i pristupačnost informacija o javnim financijama na internetskoj stranici. Promatraju i ukupnu kvalitetu internetske stranice u terminima financijskog informiranja kao skup tri spomenuta indeksa.

Kad je indeks kreiran i izmjeren, važno je utvrditi postoje li varijable koje ga određuju. Ovisno o obliku indeksa, istraživači najčešće koriste nekoliko tehnika za traženje odrednica proračunske transparentnosti: korelacijske testove, višestruku linearnu regresiju, generalizirane linearne modele te klasteriranje.

Najbliži je ovom radu onaj Laswalda, Fishera i Oyelera (vidi [22]) o odrednicama dobrovoljnog izvještavanja putem Interneta. Promatra se 86 lokalnih jedinica na Novom Zelandu. Angažman lokalnih vlasti mjeri se kao binarna varijabla koja odražava pružaju li financijske informacije na Internetu. Korištenjem logističke regresije ispituju se varijable koje bi mogle biti povezane s dobrovoljnim otkrivanjem informacija. Kao što će se ovdje promatrati sve lokalne jedinice Zadarske županije, tako su uzorak za njihovo istraživanje sve lokalne jedinice Novog Zelanda. S tim saznanjem se rezultati mogu interpretirati na drugačiji način. Naime, dok ne postoji namjera da se rezultati tog istraživanja prošire na lokalne jedinice van Novog Zelanda, uzorak istraživanja može se promatrati kao stopostotni, a kada uzorak predstavlja cijelu populaciju, testovi značajnosti općenito imaju manje smisla jer svaki regresijski koeficijent različit od nule mora biti različit od nule za populaciju u cijelosti. Dakle, svaki se regresijski koeficijent radije smatra stvarnom populacijskom vrijednošću nego procjenom populacijske vrijednosti.

Ma i Wu (vidi [24]), istražujući faktore povezane s proračunskom transparentnošću kineskih pokrajina, koriste dvoperiodnu panelnu regresijsku analizu. Po Hsiaou (vidi [17]),

panelni skup podataka prati dani uzorak pojedinaca kroz vrijeme, dakle uključuje dvije dimenzije - presječnu i vremensku. Tako pruža višestruka opažanja za svakog pojedinca u uzorku. On posjeduje više značajnih prednosti u odnosu na jednodimenzionalne modele - procjene parametara modela su točnije, a računanje i statističko zaključivanje pojednostavljeno. U ovom radu takav pristup, nažalost, nije moguć jer se potrebni podatci za moguće odrednice prerijetko skupljaju, a proračunska transparentnost mjeri tek zadnje dvije godine.

Hipoteze

U ovome se odjeljku razvijaju hipoteze o mogućim odrednicama proračunske transparentnosti lokalnih jedinica Zadarske županije. Temelje se na teorijskim pretpostavkama i empirijskim rezultatima. Proračunska se transparentnost u RH mjeri brojem objavljenih proračunskih dokumenata na internetskim stranicama lokalnih jedinica, pa se pri odabiru varijabli uzima u obzir da se radi o izvještavanju putem Interneta.

Hipoteza 1: broj stanovnika lokalne jedinice pozitivno utječe na proračunsku transparentnost

U mnogim je istraživanjima pokazan pozitivan utjecaj veličine lokalne jedinice na razinu proračunske transparentnosti, primjerice kod Guillamón, Bastide i Benita (vidi [16], 402.) koji veličinu lokalne jedinice mjere brojem stanovnika, točnije prirodnim logaritmom tog broja. Takav rezultat opravdavaju dvama razlozima. Prvo, veće lokalne jedinice primaju više javnih sredstava stoga se suočavaju s većim pritiscima. Drugo, transparentnost zahtijeva više ljudskih i materijalnih resursa što je lakše ostvarivo većim lokalnim jedinicama.

Hipoteza 2: postotak starijeg stanovništva pozitivno utječe na proračunsku transparentnost

Piotrowski i Van Ryzin (vidi [28], 310.) pokazuju da veći udio starijeg stanovništva vodi većoj potražnji za transparentnošću. To opravdavaju dokazima da godine utječu na razinu političke participacije, građanskog angažmana i povjerenja u vlast. Stariji ljudi cijene transparentnost i stoga je i zahtijevaju. Albalade del Sol (vidi [35], 97.) dodaje da neki radovi pokazuju da ta varijabla može imati suprotan učinak, ovisno o tome kako se proračunska transparentnost mjeri.

Hipoteza 3: računalna pismenost pozitivno utječe na proračunsku transparentnost

Gandía i Archidona (vidi [15]) navode da su tehnologija i obrazovanje dva najutjecajnija faktora u objašnjavanju korištenja Interneta. Stoga smatraju da je vjerojatnije da će osoba s tehnološkim pristupom i znanjem, prvenstveno znanjem korištenja Interneta, i obrazovanjem posjetiti internetsku stranicu lokalne jedinice.

Hipoteza 4: postotak kućanstava s Internetom pozitivno utječe na proračunsku transparentnost

Caba Pérez i dr. (vidi [6]) primjećuju da je ispitivanje korištenja Interneta kao potencijalne odrednice proračunske transparentnosti gotovo potpuno zanemareno i zaista pokazuju da je pristup kućanstava Internetu u pozitivnoj vezi s proračunskom transparentnošću. Debrečeny, Gray i Rahman (vidi [9]) navode da korištenje Interneta iz korisničke perspektive stvara potražnju za financijskim dokumentima, a iz perspektive vlasti stvara kanal za učinkovitije širenje informacija.

Hipoteza 5: razina obrazovanja pozitivno utječe na proračunsku transparentnost

Piotrowski i Van Ryzin (vidi [28]) smatraju da bi fakultetsko obrazovanje pojedincima moglo pružiti samopouzdanje kako bi zatražili informacije od vladajućih te potrebne vještine za razumijevanje dobivenih podataka. Analizom pokazuju kako je fakultetsko obrazovanje zaista pozitivno povezano sa zahtjevima za proračunskom transparentnošću.

Hipoteza 6: postotak muškaraca pozitivno utječe na proračunsku transparentnost

Zadarska je županija prilično tradicionalna pa bi veći postotak muškaraca u stanovništvu mogao voditi višoj transparentnosti. Žene su manje politički aktivne, i to Jennings (vidi [21]) objašnjava trima faktorima. Prvo, žene imaju zatvorene i izolirajuće uloge supruge, majke i domaćice koje im ne osiguravaju vrijeme, novac i opće vještine povezane s političkom aktivnošću. Dalje se tvrdi da ključne institucije u domenama obrazovanja, ekonomije i prava uskraćuju ženama prava i zaštitu koje imaju muškarci. Sljedeći je argument da se djecu od malih nogu uči kako je političko djelovanje odraslih više muška negoli ženska uloga. Provedeno je i empirijsko istraživanje u osam zemalja koje je potvrdilo nejednakost u političkoj participaciji između žena i muškaraca. Istina da istraživanje datira iz 1983., ali bi rezultati u ovom slučaju još uvijek mogli biti primjenjivi.

Hipoteza 7: reizbor negativno utječe na proračunsku transparentnost

Ponovni izbor lokalnog vođe može se smatrati posljedicom povjerenja u njegovo djelovanje, a Piotrowski i Van Ryzin (vidi [28]) pokazuju da takvo povjerenje negativno utječe na proračunsku transparentnost jer podrazumijeva manju potrebu za formalnim nadzorom.

Hipoteza 8: politička konkurencija pozitivno utječe na proračunsku transparentnost

Laswald, Fisher i Oyelere (vidi [22], 107.) smatraju da su uz veću političku konkurenciju izabrani političari motiviraniji održati predizborna obećanja i pružiti informacije građanima zbog većeg rizika poraza na budućim izborima. U više je radova pokazano da politička konkurencija pozitivno utječe na transparentnost, ali je mjera za političku konkurenciju različito definirana. Esteller-Moré i Polo Otero (vidi [11]) promatraju standardnu

devijaciju postotka glasova na izborima za svaku političku stranku. Što je disperzija veća, politička je konkurencija manja. Analizom pokazuju da veći stupanj političke konkurencije povećava razinu proračunske transparentnosti.

Hipoteza 9: izlaznost na izbore pozitivno utječe na proračunsku transparentnost

Veći bi odaziv na izbore mogao upućivati na to da se građani više zanimaju za aktivnosti vladajućih. U mnogim se istraživanjima ta varijabla nije pokazala značajnom, ali primjerice Caamaño-Alegre i dr. (vidi [5]) pokazuju utjecaj izborne izlaznosti na jednu od sastavnica indeksa transparentnosti, a tu su varijablu uzeli pozivajući se na članak od Esteller-Moré i Polo Otero (vidi [12]) gdje je potvrđena pozitivna veza.

Hipoteza 10: odmak od proračunske ravnoteže pozitivno utječe na proračunsku transparentnost

Alt, Lassen i Rose (vidi [2]) pokazuju da je odmak od proračunske ravnoteže povezan s većom proračunskom transparentnošću. Iako se istraživanje temeljilo na razini saveznih država SAD-a, njegovo bi objašnjenje moglo vrijediti i na lokalnoj razini: u lošim vremenima (veliki proračunski deficit) političari se trebaju opravdati i objasniti svoje postupke, a političari s dobrim rezultatima (veliki proračunski suficit) vjerojatno će se htjeti pohvaliti.

Hipoteza 11: dohodak po stanovniku pozitivno utječe na proračunsku transparentnost

Styles i Tennyson (vidi [36]) pokazuju da je varijabla dohotka po stanovniku statistički značajna u objašnjenju proračunske transparentnosti općina SAD-a. Građani s višim dohotkom više nadziru političare i traže informacije o njihovom učinku. Oni također u pravilu češće koriste Internet pa neće biti začuđujuće ako bi to značilo da će očekivati i zahtijevati proračunske dokumente u prikladnom elektroničkom obliku.

Hipoteza 12: proračunski prihodi po stanovniku pozitivno utječu na proračunsku transparentnost

Bogatstvo lokalnih jedinica također bi trebalo pozitivno utjecati na razinu proračunske transparentnosti jer se može protumačiti kao znak dobrog upravljanja pa bi političarima bilo u interesu da omoguće građanima informiranje i tako povećaju vjerojatnost ponovnog izbora. Laswald, Fisher i Oyelere (vidi [22]) uzeli su proračunske prihode po stanovniku kao mjeru komunalnog bogatstva i pokazali pozitivan utjecaj na proračunsku transparentnost.

Hipoteza 13: rashodi za računalne usluge pozitivno utječu na proračunsku transparentnost

Unatoč nedostatku empirijskih dokaza, moglo bi se pretpostaviti da lokalna jedinica koja više troši za računalne usluge i programe može lakše objavljivati potrebne informacije

na svojim internetskim stranicama, a i sam bi uzrok toj zainteresiranosti za tehnologiju mogla biti želja za komunikacijom s građanima elektroničkim putem.

Ukratko, u ovom su poglavlju navedeni radovi koji najviše doprinose temi proračunske transparentnosti, krenuvši od samog objašnjenja pojma, načina mjerenja te analize odrednica. Na temelju teorijskih i empirijskih istraživanja postavljeno je više hipoteza. Odabrane potencijalne odrednice možemo grupirati u:

- socio-kulturološke varijable: veličina lokalne jedinice, postotak starijeg stanovništva, računalna pismenost, postotak kućanstava s Internetom, razina obrazovanja, postotak muškaraca
- političke varijable: reizbor (grado)načelnika, politička konkurencija, izlaznost na izbore
- ekonomske varijable: odmak od proračunske ravnoteže, dohodak po stanovniku, proračunski prihodi po stanovniku, rashodi na računalne usluge

Pretpostavka je da reizbor negativno utječe na proračunsku transparentnost, dok sve ostale varijable imaju pozitivan utjecaj.

Poglavlje 2

Teritorijalni ustroj i proračunska transparentnost u Republici Hrvatskoj

Nakon pregleda literature stranih autora na temu proračunske transparentnosti, spušta se na razinu RH. U prvom odjeljku ovog poglavlja čitatelja se upoznaje sa sustavom lokalne i područne samouprave u RH, dok se u drugom odjeljku predstavlja proračunska transparentnost u RH kroz zakonske odredbe, način mjerenja i trenutno stanje. Tema ovog rada je analiza odrednica proračunske transparentnosti gradova i općina Zadarske županije, stoga se poseban naglasak stavlja na lokalne jedinice Zadarske županije.

2.1 Sustav lokalne i područne samouprave u Republici Hrvatskoj

Ustavom RH propisano je da je državna vlast ustrojena na načelu diobe vlasti na zakonodavnu, izvršnu i sudbenu, a ograničena je Ustavom zajamčenim pravom građanima na lokalnu i područnu (regionalnu) samoupravu.

Sadašnji sustav lokalne samouprave u RH uspostavljen je 1993., stupanjem na snagu zakona kojima se uređuje teritorijalni ustroj, samoupravni djelokrug, izborni sustav, način financiranja lokalne samouprave, a oživotvoren provedbom prvih lokalnih izbora. Jedinice lokalne samouprave su općine i gradovi, a jedinice područne samouprave su županije. Ustrojeno je ukupno 555 jedinica lokalne samouprave, 428 općina i 127 gradova, te 20 jedinica područne samouprave, odnosno županija. Grad Zagreb, kao glavni grad, ima poseban status grada i županije.

Općine i gradovi, u svom samoupravnom djelokrugu, obavljaju poslove lokalnog značaja kojima se neposredno ostvaruju potrebe građana, a koji nisu Ustavom ili zakonom dodijeljeni državnim tijelima, dok županije obavljaju poslove od područnog značaja. Za-

konom o lokalnoj i područnoj samoupravi utvrđeno je da u djelokrug općina i gradova spadaju uređenje naselja i stanovanja, prostorno i urbanističko planiranje, komunalno gospodarstvo, briga o djeci, socijalna skrb, primarna zdravstvena zaštita, odgoj i osnovno obrazovanje, kultura, tjelesna kultura i sport, zaštita potrošača, zaštita i unapređenje prirodnog okoliša, protupožarna i civilna zaštita, promet na svome području te ostali poslovi sukladno posebnim zakonom.

Jedinice lokalne i područne samouprave imaju predstavničko i izvršno tijelo koje se bira na neposrednim izborima, tajnim glasanjem, na mandat od četiri godine. Predstavnička tijela su općinsko vijeće, gradsko vijeće i županijska skupština. Izvršno tijelo u općini je općinski načelnik, u gradu gradonačelnik i u županiji župan (vidi [37]).

Svaka jedinica lokalne i područne samouprave ima svoj proračun - akt kojim se procjenjuju prihodi i primici te utvrđuju rashodi i izdaci za jednu godinu, u skladu sa zakonom i odlukom donesenom na temelju zakona, a donosi ga njezino predstavničko tijelo (vidi [30]).

Zadarsku županiju čine 34 lokalne jedinice - 6 gradova i 28 općina.

2.2 Proračunska transparentnost u Republici Hrvatskoj

Zakonske odredbe

Svi građani kao porezni obveznici imaju pravo znati kako se porezni novac troši i u što se ulaže, odnosno kako se pravi proračun RH i jedinica lokalne i područne samouprave. Čelnici lokalnih jedinica su dužni ispunjavati odredbe Zakona o pravu na pristup informacijama, Zakona o proračunu, Zakona o fiskalnoj odgovornosti i niza drugih propisa kojima je cilj informiranje javnosti o proračunu i proračunskim procesima.

Zakonom o pravu na pristup informacijama se, između ostalog, uređuje pravo na pristup informacijama koje posjeduju tijela javne vlasti i propisuju načela, ograničenja, postupak i način ostvarivanja i olakšavanja pristupa informacijama. Podatke o izvoru financiranja, proračun, financijski plan te podatke i izvješća o izvršenju proračuna jedinice lokalne i područne samouprave obvezne su objavljivati na internetskim stranicama na lako pretraživ način i u strojno čitljivom obliku (vidi [33]).

Zakon o proračunu nalaže da se proračun donosi i izvršava u skladu sa sedam proračunskih načela, u koje spada i načelo transparentnosti. To znači pravovremeno, potpuno, točno i jasno informiranje o odlukama vezanim za proračun i njihovoj provedbi. Jedinice lokalne i područne samouprave su dužne u službenom glasilu, odnosno svojim internetskim stranicama, objaviti proračun i projekciju, izmjene i dopune proračuna, odluku o privremenom financiranju te polugodišnji i godišnji izvještaj o izvršenju proračuna. Godišnje financijske izvještaje trebaju objaviti u roku od osam dana od dana njihove predaje (vidi [31]).

Mjerenje proračunske transparentnosti

Od 2013. IJF analizira i javno objavljuje rezultate mjerenja proračunske transparentnosti lokalnih i područnih jedinica. U prethodna dva ciklusa u analizi se uzimaju sve lokalne jedinice RH, za razliku od prijašnjih ciklusa kada se promatralo samo 100 općina od postojećih 428.

U ovom se radu koriste rezultati najnovije analize u kojoj IJF proračunsku transparentnost lokalnih jedinica mjeri brojem ključnih proračunskih dokumenata objavljenih na službenim internetskim stranicama lokalne jedinice u razdoblju studeni 2015.-ožujak 2016. Ti dokumenti su: godišnje izvršenje proračuna za 2014., polugodišnje izvršenje proračuna za 2015., prijedlog proračuna za 2016., izglasan proračun za 2016. i proračun za građane za 2016. Stoga se razina proračunske transparentnosti može kretati od 0 (nije objavljen niti jedan od traženih dokumenata) do 5 (objavljeni su svi traženi dokumenti).

Analiza uzima u obzir samo dokumente koji su bili dostupni u navedenom razdoblju i to na dan pretraživanja stranice. Naknadno objavljeni dokumenti se smatraju neobjavljenima, što je u potpunosti u skladu s pravovremenošću kao bitnom karakteristikom proračunske transparentnosti. Postoji mogućnost da istraživači nisu pronašli tražene dokumente, iako su bili objavljeni, ali to samo znači da bi ih i građani teško našli jer nisu bili dovoljno dobro istaknuti na internetskim stranicama (vidi [27]).

U ovom se radu za indeks proračunske transparentnosti koristi oznaka OLBI (engl. *Open Local Budget Index*).

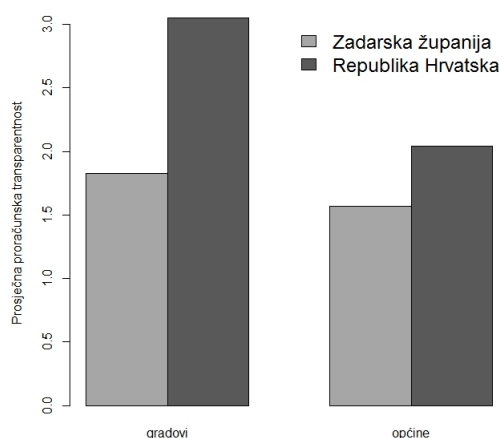
Stanje proračunske transparentnosti

Prosječna razina transparentnosti proračuna hrvatskih lokalnih i područnih jedinica je vrlo niska, iznosi tek 2.35, ali se ipak s godinama popravlja. Najbolje su županije s prosječnom transparentnošću 4.3, slijede gradovi s prosjekom 3.05, dok je prosječna transparentnost općina stvarno slaba (2.04) u čemu značajnu ulogu igra gotovo petina općina koje nisu objavile niti jedan dokument.

Iako je sama Zadarska županija postigla najvišu razinu proračunske transparentnosti, na slici 2.1 se može vidjeti da je po prosječnoj ocjeni svojih lokalnih jedinica ispod prosjeka u RH. Jedino grad Zadar ima ocjenu 5, a skoro trećina lokalnih jedinica nije objavila niti jedan od traženih proračunskih dokumenata. Stoga ne čudi sto je prosječna proračunska transparentnost gradova 1.83, a općina 1.57.

Najviše zabrinjava što je samo pet lokalnih jedinica objavilo prijedlog proračuna, a tek dvije objašnjenje za građane. Prijedlog proračuna je osobito važan jer omogućuje građanima sudjelovanje u planiranju proračuna. Građanima su proračuni preopsežni i teško razumljivi, stoga im proračunski vodič za građane omogućuje snalaženje i kreiranje cjelovite slike o proračunu.

Slika 2.1: Usporedba prosječne proračunske transparentnosti Zadarske županije i Republike Hrvatske



Izvor: autorica na temelju [27]

Ukratko, jedinice područne samouprave u RH su županije, a lokalne samouprave općine i gradovi. Zadarsku županiju čini 28 općina i 6 gradova. Svaka jedinica ima svoj proračun, a mjera razine proračunske transparentnosti je OLBI koji IJF mjeri brojem objavljenih ključnih proračunskih dokumenata na internetskim stranicama lokalne, odnosno područne jedinice. Traženi proračunski dokumenti su izabrani u skladu sa zakonima vezanim za proračun i slobodu informacija. Razina proračunske transparentnosti općenito je niska, ali se s vremenom popravlja. Zadarska županija je po prosječnoj razini proračunske transparentnosti svojih lokalnih jedinica ispod prosjeka RH.

Poglavlje 3

Metodologija

Regresijska je analiza statistički alat kojim se ispituje ovisnost jedne varijable o jednoj ili više drugih varijabli. Varijabla od interesa naziva se zavisna ili odzivna varijabla i obično označava s y , a ostale varijable koje utječu na nju nazivaju se nezavisne varijable, regresijske varijable ili prediktori i označavaju s \mathbf{x} . Model kojim se izražava statistička veza između odzivne varijable i prediktora naziva se regresijski model. Opći je oblik regresijskog modela $y = f(\mathbf{x}) + \varepsilon$ ili $y = f(\mathbf{x}) \cdot \varepsilon$. Varijabla ε izražava nepoznata odstupanja od funkcionalnog odnosa.

Ovo je poglavlje temeljeno na [25], [7], [10], [8] i [3]. U prvom se odjeljku čitatelja upoznaje s generaliziranim linearnim modelima kao generalizacijom klasičnih linearnih modela. Drugi odjeljak detaljno objašnjava prilagodbu modela podacima kroz sve procese u tom postupku, a treći odjeljak rezultate iz prva dva odjeljka primjenjuje na određeni oblik generaliziranih linearnih modela - Poissonov regresijski model.

3.1 Generalizirani linearni modeli

Klasični linearni modeli

Generalizirani linearni modeli (dalje u tekstu: GLM) proširenje su klasičnih linearnih modela. Kod linearnih se modela pretpostavlja da je vektor opažanja \mathbf{y} , koji ima n komponenti, realizacija slučajnog vektora $\mathbf{Y} = (Y_1, \dots, Y_n)$ s očekivanjem $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ čije su komponente nezavisno distribuirane slučajne varijable.

Sistematski je dio modela specifikacija vektora $\boldsymbol{\mu}$ u terminima malog broja nepoznatih parametara β_1, \dots, β_p , čije se vrijednosti procjenjuju iz podataka. Ako s i označimo indeks opažanja, tada se sistematski dio u slučaju klasičnih linearnih modela može napisati na način $\mathbb{E}(Y_i) = \mu_i = \sum_1^p x_{ij} \cdot \beta_j, i = 1, \dots, n$, gdje je x_{ij} vrijednost j -tog prediktora za

jedinku i . U matricnoj notaciji možemo zapisati $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ gdje je \mathbf{X} matrica podataka, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ vektor parametara, a $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ vektor očekivanja. U matrici podataka prikazane su mjerene vrijednosti neovisnih varijabli tako da svaki redak predstavlja jedinku istraživanja, a svaki stupac nezavisnu varijablu istraživanja. Struktura sistematskog dijela modela pretpostavlja da se kovarijate koje utječu na očekivanje znaju i da se mogu mjeriti efektivno bez grešaka.

Za slučajni dio modela pretpostavlja se nezavisnost i konstantna varijanca grešaka. Te su pretpostavke snažne i trebaju se provjeriti na podacima. Daljnje usavršavanje modela uključuje snažnije pretpostavke da greške slijede normalnu razdiobu s konstantnom varijancom σ^2 .

Klasične linearne modele stoga možemo sažeti kroz oblik - komponente od \mathbf{Y} nezavisne su normalne slučajne varijable s konstantnom varijancom σ^2 i $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu}$ gdje $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$.

Generalizacija linearnih modela

Kako bi se pojednostavnio prelazak na GLM-e, preslagivanjem gornje tvrdnje dobiva se sljedeća trodijelna specifikacija linearnih modela:

1. Slučajna komponenta: komponente od \mathbf{Y} imaju nezavisnu normalnu razdiobu s $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu}$ i konstantnom varijancom σ^2 .
2. Sistematska komponenta: kovarijate x_1, x_2, \dots, x_p stvaraju linearni prediktor η dan sa $\eta = \sum_1^p x_j \beta_j$.
3. Veza između slučajne i sistematske komponente: $\boldsymbol{\mu} = \boldsymbol{\eta}$.

Ova specifikacija uvodi novi simbol η za linearni prediktor i treću komponentu koja određuje da su $\boldsymbol{\mu}$ i $\boldsymbol{\eta}$ kod klasičnih linearnih modela zapravo jednaki. Uz zapis $\eta_i = g(\mu_i)$, funkciju $g(\cdot)$ naziva se funkcija veze (engl. *link function*).

U ovoj formulaciji, klasični linearni modeli imaju normalnu razdiobu u prvoj komponenti i identitetu za funkciju veze u trećoj komponenti. GLM-i dopuštaju dva proširenja:

- razdioba u prvoj komponenti ne mora nužno biti normalna, nego može doći iz eksponencijalne familije
- funkcija veze u trećoj komponenti može biti bilo koja monotona diferencijabilna funkcija

Iako iz teorijskih razloga slučajna komponenta odziva mora imati razdiobu iz eksponencijalne familije, to ograničenje je u praksi prihvatljivo jer eksponencijalna familija uključuje najčešće korištene razdiobe.

Funkcija vjerodostojnosti za generalizirane linearne modele

Kod GLM-a pretpostavlja se da svaka komponenta od Y ima razdiobu koja pripada eksponencijalnoj porodici, tj. gustoća joj ima oblik

$$f_Y(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\} \quad (3.1)$$

za neke funkcije $a(\cdot)$, $b(\cdot)$ i $c(\cdot)$. Porodica ima dva parametra: prirodni parametar θ i parametar disperzije ili skaliranja ϕ .

Funkcija log-vjerodostojnosti (engl. *log-likelihood*)

$$l(\theta, \phi; y) = \log f_Y(y; \theta, \phi) \quad (3.2)$$

je funkcija θ , ϕ i danog y . Očekivanje i varijanca od Y mogu se lako izvesti iz poznatih relacija iz statističke teorije:

$$\mathbb{E}\left(\frac{\partial l}{\partial \theta}\right) = 0 \quad (3.3)$$

$$\mathbb{E}\left(\frac{\partial^2 l}{\partial \theta^2}\right) + \mathbb{E}\left(\frac{\partial l}{\partial \theta}\right)^2 = 0. \quad (3.4)$$

Iz (3.2) i (3.1) slijedi

$$l(\theta; y) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \quad (3.5)$$

odakle

$$\frac{\partial l}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)} \quad (3.6)$$

i

$$\frac{\partial^2 l}{\partial \theta^2} = \frac{-b''(\theta)}{a(\phi)} \quad (3.7)$$

gdje crtice označavaju derivaciju s obzirom na θ .

Iz (3.3) i (3.6) slijedi

$$0 = \mathbb{E}\left(\frac{\partial l}{\partial \theta}\right) = \frac{\mu - b'(\theta)}{a(\phi)} \quad (3.8)$$

pa je onda

$$\mathbb{E}(Y) = \mu = b'(\theta). \quad (3.9)$$

Slično iz (3.4), (3.6) i (3.7) slijedi

$$0 = \frac{-b''(\theta)}{a(\phi)} + \frac{\text{Var}(Y)}{a^2(\phi)} \quad (3.10)$$

pa

$$\text{Var}(Y) = b''(\theta)a(\phi). \quad (3.11)$$

Stoga je varijanca od Y produkt dviju funkcija:

- $b''(\theta)$ ovisi samo o prirodnom parametru (i stoga o očekivanju) i zove se funkcija varijance. Funkcija varijance promatrana kao funkcija od μ zapisivat će se $V(\mu)$.
- $a(\phi)$ je neovisna o θ i ovisi samo o ϕ .

Funkcija $a(\phi)$ je obično oblika

$$a(\phi) = \frac{\phi}{w} \quad (3.12)$$

gdje je ϕ , koji se obilježava i sa σ^2 , konstantan kroz opažanja, a w je poznati faktor težine (engl. *prior weight*) koji se razlikuje od opažanja do opažanja.

Mnoge dobro poznate razdiobe pripadaju eksponencijalnoj familiji, a najvažnije (normalna, Poissonova, binomna i gama razdioba) su predstavljene u tablici 6.1 u Dodatku.

Funkcija veze

Funkcija veze povezuje linearni prediktor η s očekivanom vrijednosti μ za dani y . U klasičnim su linearnim modelima očekivanje i linearni prediktor jednaki, i funkcija identitete vjerodostojna je funkcija veze jer i η i μ mogu poprimiti bilo koju realnu vrijednost. Ipak, kad je riječ o prebrojivim podacima i razdioba je Poissonova, mora vrijediti $\mu > 0$, pa identiteta kao funkcija veze više nije privlačna, jer η može biti negativan dok μ ne smije. Za binomnu razdiobu vrijedi $0 < \mu < 1$ pa funkcija veze mora zadovoljavati da od cijelog brojevnog pravca poprima vrijednosti samo intervala $(0, 1)$.

Svaka od razdioba u tablici 6.1 u Dodatku ima posebnu funkciju veze za koju postoji dovoljna statistika dimenzijom jednaka β u linearnom prediktoru $\eta = \sum x_j \beta_j$. Takve funkcije veze nazivaju se prirodne ili kanonske funkcije veze i javljaju se kada $\theta = \eta$, gdje je θ prirodni parametar definiran kao u (3.1) i prikazan u tablici 6.1.

Iako su kanonske funkcije veze jednostavne i vode poželjnim statističkim svojstvima modela, osobito kod malih uzoraka, nije ih nužno uvijek koristiti.

3.2 Prilagodba modela

Proces prilagodbe modela podacima (engl. *model fitting*) može se smatrati načinom zamjene skupa y vrijednosti podataka skupom predviđenih vrijednosti $\hat{\mu}$ izvedenih iz modela koji uključuje relativno mali broj parametara.

Mogu se istaknuti četiri procesa u prilagodbi modela podacima: formulacija modela, procjena parametara, reziduali i provjera adekvatnosti modela, zaključci i interpretacija. Prilagodba se ne sastoji tek u jednoj primjeni tih procesa, već se pretpostavlja da će biti krivih pretpostavki koje će zahtijevati vraćanje na neki od prethodnih procesa u tijeku analize.

Formulacija modela

Promatraju se modeli s jednom odzivnom varijablom Y i više eksplanatornih varijabli. Model ima dvije komponente: vjerojatnosnu razdiobu od Y i jednadžbu koja povezuje očekivanu vrijednost od Y s linearnom kombinacijom eksplanatornih varijabli.

Već je rečeno da za generalizirane linearne modele vjerojatnosne razdiobe pripadaju ekspanencijskoj porodici razdioba.

Procjena parametara

Nakon odabira određenog modela, potrebno je procijeniti parametre i ocijeniti preciznost procjene. U slučaju generaliziranih linearnih modela, procjena počinje definiranjem mjere prikladnosti modela (engl. *goodness of fit*, dalje u tekstu: GOF) između opaženih podataka i predviđenih odziva generiranih modelom.

Procjene parametara vrijednosti su koje minimiziraju, odnosno maksimiziraju, određeni kriterij prikladnosti modela. U ovom radu isključivo će se objasniti procjena dobivena maksimiziranjem funkcije vjerodostojnosti, odnosno log-vjerodostojnosti.

Za dane podatke i izabranu vjerojatnosnu razdiobu, funkcija vjerodostojnosti je vjerojatnost tih podataka, tretirana kao funkcija nepoznatog parametra. Procjenitelj maksimalnom vjerodostojnošću (engl. *maximum likelihood estimator*, dalje u tekstu: MLE) vrijednost je parametra koja maksimizira funkciju vjerodostojnosti, dakle vrijednost parametra pod kojom opaženi podatci imaju najveću vjerojatnost pojavljivanja. Vrijednost parametra koja maksimizira funkciju vjerodostojnosti također maksimizira logaritam te funkcije, a log-vjerodostojnost jednostavnije je maksimizirati s obzirom na to da se tada radi o zbroju, a ne produktu članova.

Ako je $f(y; \theta)$ funkcija gustoće ili vjerojatnosna razdioba za opažanje y uz dan parametar θ , tada je log-vjerodostojnost samo

$$l(\theta; y) = \log f(y; \theta).$$

Log-vjerodostojnost temeljena na skupu nezavisnih opažanja y_1, \dots, y_n samo je zbroj pojedinačnih doprinosa, pa je

$$l(\theta; \mathbf{y}) = \sum_i \log f_i(y_i; \theta_i)$$

gdje je $\theta = (\theta_1, \dots, \theta_n)$.

Treba primijetiti da je funkcija gustoće $f(y; \theta)$ promatrana kao funkcija od y za fiksni θ dok se log-vjerodostojnost smatra prvenstveno funkcijom od θ za određeni opaženi podatak y pa odatle slijedi promjena redoslijeda argumenata.

Uz vektor nepoznatih parametara β MLE je obično lokalni maksimum koji rješava uvjete prvog reda

$$\frac{\partial l(\beta; \mathbf{y})}{\partial \beta} = \sum_{i=1}^n \frac{\partial \log f_i(y_i | \mathbf{x}_i, \beta)}{\partial \beta} \quad (3.13)$$

tj. takav $\hat{\beta}$ da

$$\left. \frac{\partial l(\beta; \mathbf{y})}{\partial \beta} \right|_{\hat{\beta}} = \mathbf{0}.$$

Za $\hat{\beta}$ ne postoji eksplicitno rješenje, već se ono dobiva iterativnim metodama (iterativni težinski najmanji kvadrati - engl. *iteratively weighted least squares*). Konvergencija je zajamčena jer je funkcija log-vjerodostojnosti globalno konkavna. U praksi je često potrebno manje od 10 iteracija.

Formalnije bi bilo da se ovaj procjenitelj naziva uvjetni MLE, jer je temeljen na uvjetnoj gustoći od y uz dani \mathbf{x} , ali uobičajeno je da se koristi jednostavniji izraz MLE.

Provjera adekvatnosti modela

Definicija ostataka

Ostatci (engl. *residuals*) mjere odmak predviđenih odziva od pravih vrijednosti zavisne varijable. Mogu se koristiti za provjeru adekvatnosti prilagodbe modela, u pogledu izbora funkcije varijance, funkcije veze i članova u linearnom prediktoru te mogu upućivati na prisutnost anomalijских podataka koje zahtijevaju daljnju istragu: stršćih i utjecajnih opažanja.

U softverskim paketima za generalizirane linearne modele najčešće se koriste dva oblika standardiziranih ostataka: Pearsonovi ostaci i ostaci u devijanci.

Obični ostatak (engl. *raw residual*) je prirodno definiran na sljedeći način

$$r_i = y_i - \hat{\mu}_i$$

gdje je $\hat{\mu}_i = \mu(\mathbf{x}_i' \boldsymbol{\beta})$ uvjetno očekivanje procijenjeno za $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$.

Pearsonovi ostaci, definirani s

$$r_p = \frac{y_i - \hat{\mu}_i}{\sqrt{\text{Var}(\hat{\mu}_i)}}$$

samo su obični ostaci skalirani procijenjenom standardnom devijacijom od Y . Ime im dolazi od činjenice da su za Poissonovu razdiobu Pearsonovi ostaci pozitivni kvadratni korijeni komponenti Pearsonove X^2 GOF statistike, tj.

$$\sum r_p^2 = X^2.$$

Ako se kao mjera odstupanja u prilagodbi generaliziranih linearnih modela koristi devijanca D , tada svaka jedinka doprinosi količinom d_i toj mjeri, pa je tako

$$\sum d_i = D.$$

Stoga ako se definira

$$r_D = \text{sign}(y_i - \mu_i) \sqrt{d_i}$$

vrijedi

$$\sum r_D^2 = D.$$

r_D se nazivaju ostaci u devijanci.

Analiza ostataka

Kad je model točan, ostaci su neovisni i imaju aproksimativno normalnu razdiobu s očekivanjem nula i konstantnom varijancom. Nadalje, ne bi trebali biti povezani s eksplanatornim varijablama. Kako bi se provjerilo jesu li pretpostavke o razdiobi opravdane, ostaci se analiziraju grafičkim testovima - histogramom i normalnim vjerojatnosnim grafom.

Najjednostavnije je nacrtati histogram relativnih frekvencija i usporediti ga s grafom funkcije gustoće pretpostavljene teoretske razdiobe, u ovom slučaju normalne. Pri tome se

nepoznati parametri procjenjuju metodom maksimalne vjerodostojnosti ili metodom momenta.

Normalni vjerojatnosni graf formira se crtanjem sortiranih podataka nasuprot kvantila očekivane vrijednosti uređene statistike standardne normalne razdiobe. Stoga se pomoću funkcije distribucije jedinične normalne razdiobe Φ definiraju brojevi $z_i = \Phi^{-1}(\frac{i-3/8}{n+1/4})$, $i = 1, \dots, n$ i u Kartezijevom koordinatnom sustavu prikazuju parovi točaka. Ako su podatci konzistentni s uzorkom iz normalne razdiobe, tada su točke aproksimativno na pravcu. Što točke variraju dalje od pravca, veća je naznaka odstupanja od normalne razdiobe.

Pearsonovi se ostatci trebaju nacrtati naspram svake eksplanatorne varijable uključene u model. Ako model adekvatno opisuje učinak varijabli, ne smije biti očitog uzorka na grafu. Sistematski uzorak na grafu ukazuje na to da se dodatni ili alternativni član možda treba uključiti u model.

Ostatci se također trebaju nacrtati naspram potencijalnih eksplanatornih varijabli koje nisu u modelu. Ako postoji ikakav sistematski uzorak na grafu, to ukazuje na to da se promatrana varijabla treba uključiti u model.

Dodatno, ostatci trebaju biti nacrtani naspram predviđenih vrijednosti, prvenstveno kako bi se otkrile promjene u varijanci. Ako se ostatci ne nalaze unutar horizontalne pruge, to ukazuje na promjenjivost varijance - heteroskedastičnost.

Konačno, potrebno je prikazati graf ostataka koristeći poredak u kojem su mjerene vrijednosti odzivne varijable. To može biti vremenski, prostorni ili bilo koji drugi slijed koji bi mogao uzrokovati ovisnost među opažanjima. Ako su opažanja neovisna, točke bi trebale fluktuirati slučajno bez ikakvog sistematskog uzorka, kao što je alterniranje prema gore ili dolje ili postojano povećavanje ili smanjenje odmak od nule.

Stršeca i utjecajna opažanja

Identifikacija neobičnih opažanja važan je dio svake regresijske analize.

Stršeca vrijednost (engl. *outlier*) je opažanje s velikom apsolutnom vrijednosti ostatka, tj. odzivna je vrijednost tog opažanja neočekivana uz dane prediktorske varijable. Široko je primijećeno da su Pearsonovi ostatci izrazito asimetrični (engl. *skewed*) u mnogim GLM primjenama pa se zbog toga kod dijagnostike preferira korištenje ostataka u devijanci kako se outlieri ne bi pogrešno identificirali.

Dijagonalni elementi "kapa" matrice (engl. *hat matrix*) h_{ii} predstavljaju mogući učinak i -tog opažanja na vlastitu predviđenu vrijednost i koriste se kao mjere za snagu (engl. *leverage*). Snažna točka (engl. *leverage point*) ima veliku vrijednost h_{ii} . Takve točke vuku

podešenu regresiju prema sebi i tako potencijalno deformiraju prilagođeni model. Također mogu imati snažan učinak na mjere jakosti opažene regresijske veze.

Utjecajno (engl. *influential*) opažanje objedinjuje veliko odudaranje (engl. *outlyingness*) i veliku snagu. Očituje se u utjecaju na regresijske koeficijente, u smislu da se uklanjanjem tog opažanja koeficijenti značajno mijenjaju. Kao mjera utjecajnosti za pojedino opažanje koristi se Cookova udaljenost (engl. *Cook's distance*). U slučaju da je Cookova udaljenost veća od jedinice, predlaže se daljnja provjera tog opažanja.

Ova je tema manje važna u primjenama u kojima su skupovi podataka veliki i relativna važnost individualnog opažanja mala. Ako su skupovi podataka mali i utjecajno opažanje otkriveno, nije uvijek jasno kako treba nastaviti. Na početku je svakako važno odrediti je li postojanje utjecajnog opažanja ili outliera uzrokovano pogreškom u mjerenju, greškom u prepisivanju ili nekim drugim propustom. Takvo se opažanje može ukloniti iz skupa podataka samo ako postoji argumentima potkrijepljen razlog. O takvim opažanjima bi se svakako trebalo izvijestiti, a moguće je rješenje zadržati to opažanje, i prijaviti rezultate koji su dobiveni s i bez uključenja tog opažanja u model.

Odabir modela i zaključci

Kod odabira modela važno je odabrati kovarijate koje će se uključiti u sistematski dio modela. Najjednostavnije rečeno, dane su određene kovarijate x_1, \dots, x_p od kojih treba naći podskup koji na neki način najbolje konstruira fitane vrijednosti.

Uz danih n opažanja može se fitati model koji sadrži do n parametara. Najjednostavniji model (engl. *null model*) ima jedan parametar, koji predstavlja zajednički μ za sve y -e - takav model svu varijaciju među y -ima pripisuje slučajnoj komponenti. Takav je model u praksi obično prejednostavan.

Druga je krajnost zasićeni (engl. *saturated, full*) model koji ima n parametara, jedan po opažanju i μ -ovi izvedeni iz njega točno odgovaraju podacima. Puni model dakle svu varijaciju u y -ima pripisuje sistematskoj komponenti ne ostavljajući ništa slučajnoj.

Na prvi pogled može se činiti da je dobar model onaj koji fita opažene vrijednosti vrlo dobro, tj. stvara $\hat{\mu}$ vrlo blizak y . Ipak zasićeni model podatke ne sažima već ih samo sve ponavlja. Jednostavnost predstavljena štedljivošću (engl. *parsimony*) u parametrima poželjna je u svakom modelu - stoga se parametri koji nisu potrebni neće uključiti u model. Ako se model vrlo dobro prilagođava određenom skupu podataka, neće biti sposoban obuhvatiti promjene do kojih će nužno doći pri prikupljanju drugog skupa podataka koji se odnosi na istu pojavu. Ipak zasićeni model daje osnovu za mjerenje odstupanja srednjeg modela s k parametara, $k < p$.

Unatoč mnogim smjernicama, modeliranje u znanosti barem djelomično ostaje umjetnost. Prvo, svaki model je netočan, ali neki modeli su korisniji od drugih i njima treba težiti. Nadalje, često postoji više mogućih modela koji dobro odgovaraju podacima. Očekuje se da će grupiran oko "najboljeg" modela biti skup gotovo jednako dobrih modela koji se statistički bitno ne razlikuju. Nadalje, redoslijed kojim se nove komponente uvode u model također je bitan. Ponekad je moguće da drugi redoslijed da drugačije rezultate, tj. drugačiji konačni model.

Prikladnost modela

Vjerodostojnost je uobičajeno izraziti u terminima parametra srednje vrijednosti μ radije nego prirodnog parametra θ . Neka je $L(\hat{\mu}, \phi; \mathbf{y})$ vjerodostojnost maksimizirana po β za fiksnu vrijednost parametra disperzije ϕ . Maksimalno ostvariva vjerodostojnost postiže se u punom modelu s n parametara i označava $L(\mathbf{y}, \phi; \mathbf{y})$.

Tada omjer vjerodostojnosti, u oznaci λ ,

$$\lambda = \frac{L(\mathbf{y}, \phi; \mathbf{y})}{L(\hat{\mu}, \phi; \mathbf{y})},$$

daje procjenu prikladnosti modela. U praksi se koristi logaritam omjera vjerodostojnosti, koji je razlika funkcija log-vjerodostojnosti

$$\log \lambda = l(\mathbf{y}, \phi; \mathbf{y}) - l(\hat{\mu}, \phi; \mathbf{y}).$$

Velika vrijednost $\log \lambda$ sugerira da promatrani model loše opisuje podatke u odnosu na pun model. Pokazuje se da izraz $2 \log \lambda$ ima χ^2 razdiobu s $n - k$ stupnjeva slobode, gdje je k broj parametara promatranog modela, stoga je $2 \log \lambda$ češće korištena statistika, i naziva se devijanca.

Oblici devijanci za razdiobe dane u tablici 6.1 u Dodatku se nalaze u tablici 6.2, također u Dodatku.

Druga je važna mjera odstupanja od prilagodbe generalizirana Pearsonova X^2 statistika, oblika

$$X^2 = \sum (y - \hat{\mu})^2 / V(\hat{\mu})$$

gdje je $V(\hat{\mu})$ procijenjena funkcija varijance za dotičnu razdiobu. Pearsonova X^2 statistika također ima asimptotski χ^2 razdiobu.

Usporedba ugniježđenih modela

Za usporedbu koliko su dva ugniježđena modela dobro prilagođena podacima, koristimo test omjera vjerodostojnosti (engl. *likelihood ratio test*). Za generalizirane linearne modele, dva modela trebaju imati jednaku vjerojatnosnu razdiobu i jednaku funkciju veze,

ali linearna komponenta jednog modela ima više parametara od drugog. Ako je jednostavniji model M_0 s q parametara, koji odgovara hipotezi H_0 , prilagođen podacima jednako dobro kao i općenitiji model M_1 s p parametara, tada mu se daje prednost na osnovu principa štedljivosti u parametrima i ne odbacuje se H_0 . Ako je općenitiji model značajno bolje prilagođen podacima, tada se H_0 odbacuje u korist alternativne hipoteze H_1 , koja se odnosi na općenitiji model. H_0 se može testirati protiv H_1 koristeći statistiku omjera vjerodostojnosti, odnosno statistiku razlika devijanci

$$\begin{aligned}\Delta D &= D_0 - D_1 = \\ &= 2[l(\mathbf{y}; \mathbf{y}) - l(\hat{\boldsymbol{\mu}}_0; \mathbf{y})] - 2[l(\mathbf{y}; \mathbf{y}) - l(\hat{\boldsymbol{\mu}}_1; \mathbf{y})] = \\ &= 2[l(\hat{\boldsymbol{\mu}}_1; \mathbf{y}) - l(\hat{\boldsymbol{\mu}}_0; \mathbf{y})].\end{aligned}\tag{3.14}$$

Ako oba modela dobro opisuju podatke, tada je $D_0 \sim \chi^2(n - q)$ i $D_1 \sim \chi^2(n - p)$ pa je onda $\Delta D \sim \chi^2(p - q)$, uz ispunjene određene uvjete nezavisnosti. Ako je vrijednost ΔD u skladu s $\chi^2(p - q)$ razdiobom, općenito se odabire model M_0 koji odgovara H_0 jer je jednostavniji. Test je egzaktn samo za normalnu razdiobu, a inače vrijedi asimptotski. Ako model M_0 ne opisuje dobro podatke, tada će D_0 biti veća nego što se očekuje za vrijednost iz $\chi^2(n - q)$. Ako model M_1 dobro opisuje podatke tako da je $D_1 \sim \chi^2(n - p)$, a M_0 ne opisuje podatke dobro, tada će ΔD biti veća od očekivane iz $\chi^2(p - q)$.

Dobiveni rezultat koristi se za testiranje hipoteze H_1 na sljedeći način: ako je vrijednost ΔD u kritičnom području, H_0 se odbacuje u korist H_1 na osnovu toga da model M_1 pruža značajno bolji opis podataka.

Uzoračka razdioba od ΔD obično je bolje aproksimirana χ^2 razdiobom nego što je to uzoračka razdioba za pojedinačnu devijancu.

Općenita usporedba modela: Informacijski kriteriji

Dva su modela neugniježđena ako se niti jedan od njih ne može prikazati kao poseban slučaj drugoga. Za usporedbu takvih modela, baziranih na maksimalnoj vjerodostojnosti, predlažu se kriteriji odabira modela temeljeni na funkciji log-vjerodostojnosti. S obzirom na to da se očekuje da se log-vjerodostojnost povećava dodavanjem parametara u model, ti kriteriji penaliziraju modele se većim k , brojem parametara u modelu. Ta funkcija penalizacije može biti i funkcija od n , broja opažanja.

Akaike informacijski kriterij (engl. *Akaike information criteria*, dalje u tekstu: AIC) definiran je s

$$AIC = -2 \log L + 2k.$$

Broj je koristan za usporedbu modela, ali sam po sebi nije značajan. Poželjan je model onaj s najnižom vrijednosti AIC kriterija. Poznato je da AIC ima sklonost voditi prekomjernom

broju parametara (engl. *overfitting*), osobito kod malih uzoraka, tj. kazneni član u AIC-u dizajniran da čuva od prekomplikiranih modela nije dovoljno snažan.

Izmijenjene inačice AIC-a pomažu se nositi s tim problemom, a najkorištenija je Bayesian informacijski kriterij (engl. *Bayesian information criteria*, dalje u tekstu: BIC) koja vrijednost 2 zamjenjuje prirodnim logaritmom veličine uzorka i tako strože kažnjava dodatne parametre:

$$BIC = -2 \log L + (\log n)k.$$

Stoga, u usporedbi s AIC-om, BIC sporije nagine kompleksnijim modelima kako se n povećava.

Automatizirane procedure odabira modela

U softverskim paketima postoje automatizirane procedure odabira najboljeg modela zasnovane na AIC-u i sličnim kriterijima.

Regresija najboljeg podskupa (engl. *best subset regression*) odabire najbolji model od svih mogućih podskupova prema nekom kriteriju prikladnosti modela. Ova metoda uključuje ispitivanje svih modela generiranih od svih mogućih kombinacija prediktorskih varijabli i odabiranje najboljeg modela svake veličine. Za p mogućih prediktorskih varijabli postoji 2^p mogućih podmodela. Javljaju se statistički problemi i problemi u izračunavanju. Samo računanje, koje ignorira bilo kakvu vezu koja može postojati među kovarijatama, može se opisati na sljedeći način: naći najboljih s podskupova veličine r među kovarijatama. Ako je broj raspoloživih kovarijata malen, recimo $p \leq 12$, najbolji podskupovi za svaki r , $r = 1, \dots, p$ mogu biti nađeni potpunim prebrojavanjem. Za veće p izvode se prečice.

Aproksimativne metode za generiranje jednog "najboljeg" modela su sljedeće:

- odabir unaprijed (engl. *forward selection*) - u svakoj fazi najbolja neizabrana kovarijata po određenom selekcijskom kriteriju dodaje se u model dok ne ponestane daljnjih kandidata
- eliminacija unatrag (engl. *backward elimination*) - počinje s punim skupom i eliminira najgoru kovarijatu jednu po jednu sve dok preostale kovarijate nisu sve potrebne
- regresija korak po korak (engl. *stepwise regression*) - kombinira prethodne dvije procedure, eliminaciju unatrag slijedi selekcija unaprijed sve dok nijedna ne mijenja model

Procedure odabira varijabli treba koristiti s oprezom. Automatizirane procedure ne smiju biti zamjena za pažljivo razmatranje modela.

3.3 Poissonov regresijski model

U mnogim primjenama varijabla od interesa je nenegativni cijeli broj. Za razliku od klasičnog linearnog regresijskog modela, odzivna je varijabla tada diskretna s razdiobom čija je gustoća koncentrirana samo na nenegativnim cijelim brojevima. Poissonov regresijski model polazna je točka za rad s ovakvim podacima, a pokušaji prilagodbe brojnim uvjetima iz stvarnog života vode do dodatnih razrađivanja modela.

Poissonova razdioba

Najjednostavnija je razdioba za modeliranje prebrojivih podataka (engl. *count data*) Poissonova razdioba, stoga je na početku korisno predstaviti neka njena ključna svojstva.

Ako je diskretna slučajna varijabla Y iz Poissonove razdiobe s parametrom intenziteta μ , $\mu > 0$, i izlaganjem t , koje je definirano kao vremenski period unutar kojeg se događaji bilježe, tada Y ima gustoću

$$\Pr(Y = y) = \frac{e^{-\mu t} (\mu t)^y}{y!}, y = 0, 1, 2, \dots \quad (3.15)$$

gdje $\mathbb{E}(Y) = \text{Var}(Y) = \mu t$.

Ako se period izlaganja t u (3.15) postavi na jedinični, tada vrijedi

$$\Pr(Y = y) = \frac{e^{-\mu} (\mu)^y}{y!}, y = 0, 1, 2, \dots \quad (3.16)$$

tj. Poissonova razdioba ima samo jedan parametar μ i označavamo je sa $P(\mu)$.

Jednakost očekivanja i varijance naziva se Poissonovo svojstvo jednake disperzije (engl. *equidispersion*). U stvarnom životu to je svojstvo često narušeno. Prevelika disperzija ili preraspršenost (engl. *overdispersion*) znači da varijanca premašuje očekivanje, a premala disperzija (engl. *underdispersion*) obrnuto.

Ključno je svojstvo Poissonove razdiobe aditivnost, preciznije ako su $Y_i \sim P(\mu_i)$, $i = 1, 2, \dots$ nezavisne slučajne varijable i $\sum \mu_i < \infty$, tada je $S_Y = \sum Y_i \sim P(\sum \mu_i)$.

Slika 3.1 pokazuje kako se oblik Poissonove razdiobe mijenja s parametrom μ . Ilustrira stupanj asimetrije, osobito za male vrijednosti μ , i približavanje normalnosti za velike vrijednosti μ . Normalnost se tu odnosi na kumulativnu razdiobu, a ne na gustoću.

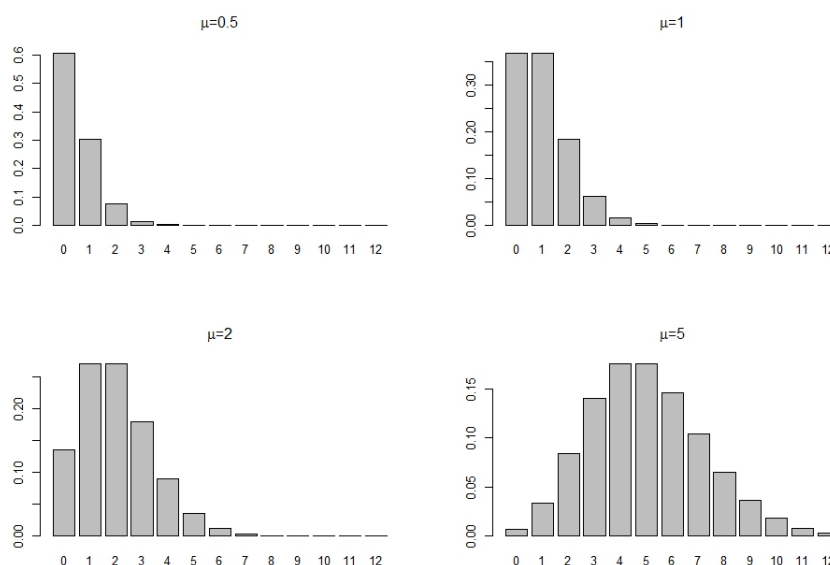
Poissonova regresija

Funkcija gustoće Poissonove razdiobe (3.16) može se zapisati u obliku

$$f(y, \mu) = \exp\{y \log \mu - \mu - \log(y!)\}.$$

Uz $\theta = \log(\mu)$, $\phi = 1$, $a(\phi) = 1$, $b(\theta) = e^\theta$, $c(y, \phi) = -\log(y!)$ lako se provjeri da Poissonova razdioba pripada eksponencijalnoj porodici, dakle Poissonova je regresija poseban slučaj u okviru generaliziranih linearnih modela.

Slika 3.1: Poissonova razdioba za $\mu = 0.5, 1, 2$ i 5



Izvor: autorica prema [25]

Učinak eksplanatornih varijabli na odziv Y modeliran je kroz parametar μ . Potrebno je naći funkciju veze $g(\cdot)$ koja ograničava μ na pozitivne vrijednosti, a linearnom prediktoru dopušta da poprimi bilo koju realnu vrijednost. Koristi se prirodni logaritam koji je kanonska funkcija veze, već navedena u tablici 6.1 u Dodatku.

Zavisna varijabla, broj pojava događaja od interesa, označena je s y_i , a vektor linearno nezavisnih prediktora za koje se smatra da određuju y_i s \mathbf{x}_i . Dakle, y_i je, uz dan \mathbf{x}_i , iz Poissonove razdiobe s gustoćom

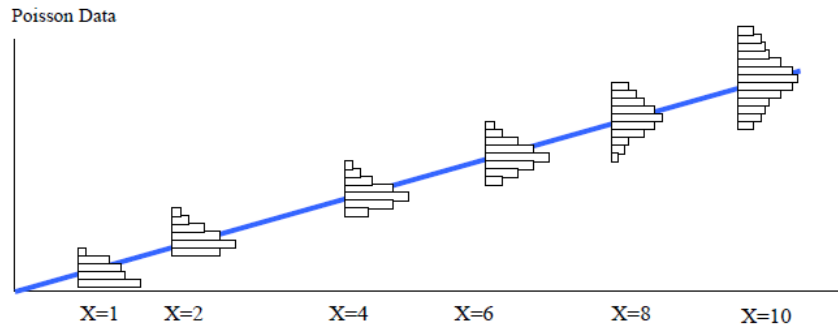
$$f(y_i|\mathbf{x}_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, y_i = 0, 1, \dots, n \quad (3.17)$$

a parametar očekivanja parametriziran je na sljedeći način

$$\mu_i = \exp(\mathbf{x}_i' \boldsymbol{\beta}). \quad (3.18)$$

Jednadžbe (3.17) i (3.18) zajedno definiraju Poissonov ili log-linearni regresijski model. Model se naziva log-linearnim jer je logaritam uvjetnog očekivanja linearan u parametrima: $\log \mathbb{E}(y_i|\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}$.

Slika 3.2: Ilustracija Poissonove regresije



Izvor: preuzeto iz [23]

Poissonovo svojstvo $\mathbb{E}(y_i|\mathbf{x}_i) = \text{Var}(y_i|\mathbf{x}_i)$ i (3.18) impliciraju da uvjetna varijanca nije konstantna, stoga je regresija intrinzično heteroskedastična.

U log-linearnoj verziji modela gdje je parametar očekivanja parametriziran prema (3.18) uvjetno očekivanje ima multiplikativan oblik dan s

$$\mathbb{E}(y_i|\mathbf{x}_i) = \exp(\mathbf{x}_i'\boldsymbol{\beta}) = \exp(x_{1i}\beta_1) \exp(x_{2i}\beta_2) \cdots \exp(x_{ki}\beta_k)$$

gdje zanimanje često leži u promjenama uvjetnog očekivanja uslijed promjena u regresorima.

Procjena parametara

Poissonov MLE $\hat{\boldsymbol{\beta}}$ maksimizira funkciju log-vjerodostojnosti

$$l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta}) = \log \prod_{i=1}^n f(y_i|\mathbf{x}_i, \boldsymbol{\beta}) = \sum_{i=1}^n \log f(y_i|\mathbf{x}_i, \boldsymbol{\beta}).$$

Za Poissonovu razdiobu, log-vjerodostojnost za opažanje i je

$$\begin{aligned}\log f(y_i|\mathbf{x}_i, \boldsymbol{\beta}) &= \log \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} = \\ &= \log(e^{-\mu_i} \mu_i^{y_i}) - \log(y_i!) = \\ &= -\mu_i + y_i \log(\mu_i) - \log(y_i!) = \\ &= -\exp(\mathbf{x}_i' \boldsymbol{\beta}) + y_i \mathbf{x}_i' \boldsymbol{\beta} - \log(y_i!)\end{aligned}\quad (3.19)$$

stoga Poissonov MLE maksimizira funkciju

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n \log f(y_i|\mathbf{x}_i, \boldsymbol{\beta}) = \sum_{i=1}^n \{-\exp(\mathbf{x}_i' \boldsymbol{\beta}) + y_i \mathbf{x}_i' \boldsymbol{\beta} - \log(y_i!)\}. \quad (3.20)$$

Derivacija po $\boldsymbol{\beta}$ daje Poissonov MLE $\hat{\boldsymbol{\beta}}$ kao rješenje uvjeta prvog reda

$$\sum_{i=1}^n (y_i - \exp(\mathbf{x}_i' \boldsymbol{\beta})) \mathbf{x}_i = \mathbf{0}.$$

Interpretacija regresijskih koeficijenata

Važno je pitanje koje se javlja u analizi interpretacija regresijskih koeficijenata. Često je zanimanje za granične efekte - promjene u uvjetnom očekivanju od y kada se regresori \mathbf{x} promjene za jednu jedinicu.

Za linearne regresijske modele, $\mathbb{E}(y|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$ implicira $\partial \mathbb{E}(y|\mathbf{x})/\partial x_j = \beta_j$ tako da koeficijenti imaju izravnu interpretaciju - jedinična promjena u j -tom regresoru povećava uvjetno očekivanje za $\hat{\beta}_j$ jedinica. Za genelinearne se linearne modele ova interpretacija mora se prilagoditi.

Promatrajući eksponencijalno uvjetno očekivanje

$$\mathbb{E}(y|\mathbf{x}) = \exp(\mathbf{x}'\boldsymbol{\beta}) \quad (3.21)$$

razmatra se učinak promjene u x_j , j -tom regresoru. Derivacija po x_j daje granični efekt jedinične promjene u x_j :

$$\frac{\partial \mathbb{E}(y|\mathbf{x})}{\partial x_j} = \beta_j \exp(\mathbf{x}'\boldsymbol{\beta}) \quad (3.22)$$

Izračunate se vrijednosti razlikuju između pojedinačnih opažanja zbog različitih vrijednosti od \mathbf{x} , što otežava interpretaciju. Zbog pojednostavljenja se računaju samo neki granični efekti: prosjek granični efekata svih jedinki, granični efekt za jedinku s prosječnim svojstvima ili granični efekt za izabrane vrijednosti \mathbf{x} od posebnog zanimanja.

Dodatna se komplikacija pojavljuje u prisutnosti binarnih regresora. To su često varijable tretmana (engl. *treatment variable*), regresor koji je jednak 1 ako je pojedinac tretiran, a 0 ako nije. Zanimanje leži u učinku tretmana (engl. *treatment effect*) na uvjetno očekivanje kad se indikatorska varijabla, nazovimo je d , promijeni iz 0 u 1. Neka je $\mathbf{x} = [d \quad \mathbf{z}]$, gdje \mathbf{z} označava sve regresore osim binarnog regresora d . Tada je efekt tretmana definiran s

$$\mathbb{E}(y|d = 1, \mathbf{z}) - \mathbb{E}(y|d = 0, \mathbf{z}) = \exp(\beta_d + \mathbf{z}'\beta_z) - \exp(\mathbf{z}'\beta_z). \quad (3.23)$$

Efekt tretmana varira s \mathbf{z} . Stoga bismo mogli definirati prosječni efekt tretmana i prosječni efekt tretmana nad tretiranima.

Korisno je primijetiti da je direktna interpretacija koeficijenata moguća bez većeg dodatnog računanja:

- Za modele s eksponencijalnim uvjetnim očekivanjem, koeficijent β_j jednak je proporcionalnoj promjeni uvjetnog očekivanja ako se j -ti regresor promjeni za jedinični iznos. To slijedi iz (3.21) i (3.22):

$$\beta_j = \frac{\partial \mathbb{E}(y|\mathbf{x})}{\partial x_j} \frac{1}{\mathbb{E}(y|\mathbf{x})}.$$

- Predznak odziva $\partial \mathbb{E}(y|\mathbf{x})/\partial x_j$ dan je predznakom od β_j jer je odziv $\beta_j \cdot \exp(\mathbf{x}'\boldsymbol{\beta})$, a $\exp(\mathbf{x}'\boldsymbol{\beta})$ je uvijek pozitivno.
- Ako je jedan regresijski koeficijent duplo veći od drugog, tada je učinak jedinične promjene pripadnog regresora dupli nego od drugog. Rezultat slijedi iz

$$\frac{\partial \mathbb{E}(y|\mathbf{x})/\partial x_j}{\partial \mathbb{E}(y|\mathbf{x})/\partial x_k} = \frac{\beta_j \exp(\mathbf{x}'\boldsymbol{\beta})}{\beta_k \exp(\mathbf{x}'\boldsymbol{\beta})} = \frac{\beta_j}{\beta_k}.$$

- Za binarni regresor efekt tretmana, dan u (3.23), može biti interpretiran i u relativnim terminima kao

$$\frac{\mathbb{E}(y|d = 1, \mathbf{z})}{\mathbb{E}(y|d = 0, \mathbf{z})} = \frac{\exp(\beta_d + \mathbf{z}'\beta_z)}{\exp(\mathbf{z}'\beta_z)} = \exp(\beta_d).$$

Stoga je uvjetno očekivanje $\exp(\beta_d)$ puta veće ako je indikatorska varijabla jedinica, a ne nula.

- Općenitije, za jediničnu promjenu u bilo kojem regresoru x_j uvjetno je očekivanje $\exp(\beta_j)$ puta veće, pa neki programi prikazuju eksponencirane koeficijente.

Preraspršenost

U praksi se kod prebrojivih opažanja javlja varijabilnost koja premašuje onu predviđenu Poissonom. U tom slučaju, kad je uvjetna varijanca veća od uvjetnog očekivanja, za podatke kažemo da su preraspršeni. Preraspršenost je blisko povezana s prisustvom neopažene heterogenosti među pojedincima, ali može proizaći i iz ovisnosti pojava među događajima. Preraspršenost može uzrokovati podcjenjivanje standardnih grešaka pa stoga i rezultati testova značajnosti mogu biti precijenjeni što vodi krivim zaključcima.

Najlakše je rješenje u prisutstvu preraspršenosti gledanje Poissonovog MLE kao kvazi MLE (dalje u tekstu: QMLE) i temeljenje zaključaka na robusnim standardnim greškama. Procjena kvazi vjerodostojnošću pretpostavlja samo vezu između očekivanja i varijance umjesto određene razdiobe za Y_i - ima funkciju veze i linearni prediktor uobičajenog GLM oblika, ali za Y_i pretpostavlja samo $\text{Var}(Y_i) = \nu(\mu_i)$ za odabranu funkciju varijance ν . U alternativni Poissonovog GLM-a funkcija veza između očekivanja i varijance dana je s $\nu(\mu_i) = \phi\mu_i$ za neku konstantu ϕ . Slučaj $\phi > 1$ predstavlja preraspršenost za Poissonov model. Jednadžbe koje određuju procjene kvazi vjerodostojnošću jednake su jednadžbama vjerodostojnosti za GLM, stoga će procjene koeficijenata biti jednake MLE u Poissonovom modelu, ali su standardne greške prilagođene za procijenjeni iznos preraspršenosti.

Dvije metode koje vode efikasnijoj procjeni samo će se kratko spomenuti. Prva je određivanje prikladnijeg parametarskog modela koji objašnjava preraspršenost. Standardni model koji se prilagođava preraspršenosti je negativni binomijalni model. Treba napomenuti da se negativni binomijalni model može prilagoditi jedino preraspršenosti. Druga metoda je manje parametarska koja procjenu temelji na određivanju prva dva momenta.

Ukratko, generalizirani linearni modeli određeni su trima komponentama: slučajnom, sistematskom i vezom između njih. Poopćuju linearne modele dopuštanjem da razdioba u slučajnoj komponenti dolazi iz bilo koje razdiobe iz eksponencijalne familije razdioba te da funkcija veze može biti bilo koja monotona diferencijabilna funkcija. Mnoge dobro poznate i često korištene razdiobe pripadaju eksponencijalnim familijama, stoga GLM-i imaju široku primjenu.

Prilagodba GLM-a podacima može se podijeliti u četiri procesa:

- Formulacija modela obuhvaća odabir vjerojatnosne razdiobe odzivne varijable i jednadžbe koja povezuje odzivnu varijablu i linearni prediktor.

- Parametri modela vrijednosti su koje maksimiziraju funkciju vjerodostojnosti, odnosno log-vjerodostojnosti. Za takav procjenitelj ne postoji eksplicitno rješenje, već se ono dobiva iterativnim metodama.
- Provjera adekvatnosti modela podrazumijeva analizu ostataka te ispitivanje postojanja neobičnih opažanja, outliera i utjecajnih opažanja, koji bi se, ako postoje, trebali bolje istražiti. Analiza ostataka može pokazati da odabrani model nije prikladan, u smislu da je loše izabrana funkcija veze ili funkcija varijance, ili da su neki važni prediktori izostavljeni iz modela.
- Za promatrani model potrebno je odrediti je li uopće prikladan za dane podatke. Usporedba prikladnosti modela podacima za dva ugniježđena modela provodi se testom omjera vjerodostojnosti, a za općenite usporedbe modela koriste se informacijski kriteriji. Tako se odabiru kovarijate koje treba uključiti u sistematski dio modela kako bi se dobio "najbolji" model, ili skup takvih.

Poissonova je varijabla diskretna s razdiobom čija je gustoća koncentrirana na nenegativnim cijelim brojevima pa je zato prikladna za modeliranje prebrojivih podataka. Poissonova razdioba pripada eksponencijalnoj familiji razdioba, stoga Poissonov regresijski model pripada generaliziranim linearnim modelima. Kao funkcija veze koristi se prirodni logaritam, odnosno uvjetno je očekivanje eksponencijalno. U praksi je kod prebrojivih podataka često narušen uvjet jednakosti uvjetnog očekivanja i varijance pa postoji više poopcjenja Poissonove regresije.

Poglavlje 4

Podatci

Tema je ovog rada analiza varijabli koje bi mogle određivati razinu proračunske transparentnosti lokalnih jedinica u Zadarskoj županiji. U Poissonovom regresijskom modelu zavisna je varijabla proračunska transparentnost, a nezavisne varijable su odabrane na temelju hipoteza postavljenih u odjeljku 1.2.

4.1 Zavisna varijabla

Zavisna je varijabla u ovom radu indeks proračunske transparentnosti OLBI opisan u odjeljku 2.2. Posljednji dostupni podatci su iz 2016. Važno je podsjetiti se da se OLBI mjeri brojem objavljenih od pet preporučenih proračunskih dokumenata na internetskoj stranici lokalne jedinice, stoga je varijabla OLBI diskretna i poprima vrijednosti od 0 do 5.

U tablici 4.1 se nalazi prikaz razine OLBI-ja gradova i općina u Zadarskoj županiji po vrsti i broju objavljenih dokumenata.

4.2 Nezavisne varijable

Nezavisne varijable koje ću ispitati grupirane su u tri glavne kategorije: socio-kulturološke, političke i ekonomske.

Podatci za socio-kulturološke varijable su s posljednjeg popisa stanovništva provedenog 2011., a rezultati su objavljeni tijekom 2012. i 2013. Za istraživanje bi nesumnjivo bilo bolje kad bi postojali noviji podatci, ali s obzirom na to da oni ne postoje, pretpostavit ćemo da se nisu znatno mijenjali ili da su se mijenjali istim intenzitetom u svim promatranim lokalnim jedinicama.

Tablica 4.1: Prikaz razine proračunske transparentnosti lokalnih jedinica u Zadarskoj županiji

Lokalna jedinica	Godišnji izvještaj o izvršenju proračuna za 2014.	Polugodišnji izvještaj o izvršenju proračuna za 2015.	Prijedlog proračuna za 2016.	Izglasani proračun za 2016.	Proračun za građane za 2016.	OLBI 2016
Gradovi						
Benkovac	0	1	0	1	0	2
Biograd na Moru	1	0	0	1	0	2
Nin	0	0	0	0	0	0
Obrovac	0	0	0	1	0	1
Pag	1	0	0	0	0	1
Zadar	1	1	1	1	1	5
Općine						
Bibinje	0	0	0	0	0	0
Galovac	0	1	0	0	0	1
Gračac	1	1	1	1	0	4
Jasenice	0	0	0	0	0	0
Kali	1	1	1	1	0	4
Kolan	0	0	0	0	0	0
Kukljica	0	0	0	1	0	1
Lišane Ostrovičke	0	0	0	0	0	0
Novigrad	0	1	0	1	0	2
Pakoštane	0	0	0	1	0	1
Pašman	0	0	0	0	0	0
Polača	1	0	0	1	0	2
Poličnik	1	0	0	1	0	2
Posedarje	0	1	0	0	0	1
Povljana	0	0	0	0	0	0
Preko	0	1	0	1	0	2
Privlaka	1	1	1	1	0	4
Ražanac	0	0	0	1	0	1
Sali	1	0	0	1	0	2
Stankovci	0	0	0	0	0	0
Starigrad	1	1	0	1	1	4
Sukošan	1	1	1	1	0	4
Sveti Filip i Jakov	1	1	0	1	0	3
Škabrnja	0	0	0	0	0	0
Tkon	0	0	0	1	0	1
Vir	1	1	0	1	0	3
Vrsi	0	0	0	0	0	0
Zemunik Donji	1	0	0	1	0	2
Ukupno	14	13	5	21	2	

Izvor: autorica prema [18]

Analiziraju se četiri ekonomske varijable. Jedna od njih je odmak ostvarenog proračuna od proračunske ravnoteže, dakle višak prihoda nad rashodima, odnosno apsolutna vrijednost iznosa ako su rashodi veći od prihoda.

Promatra se i prosječni dohodak po stanovniku, za razdoblje od 2010. do 2012. Dohodak po stanovniku izračunava se kao omjer ukupnog iznosa dohotka kojega su tijekom jednoga poreznog razdoblja (kalendarska godina) ostvarili porezni obveznici, fizičke osobe s prebivalištem ili uobičajenim boravištem na području jedinice lokalne, odnosno područne samouprave za koju se vrši izračun, i broja stanovnika koji žive na području te jedinice (vidi [32]). Sljedeći bi izračun trebao biti objavljen 2018., stoga se trenutno mora zadovoljiti ovim podacima.

Tablica 4.2: Opis nezavisnih varijabli i izvori podataka

Naziv varijable	Opis varijable	Razdoblje	Izvor podataka
broj.st	broj stanovnika	2011.	DZS [42]
visokoobraz	postotak visokoobrazovanih u stanovništvu starijem od 20 godina (stručni studij, sveučilišni studij ili doktorat znanosti)	2011.	DZS [42], [41]
muskarci	postotak muškaraca u ukupnom stanovništvu	2011.	DZS [42]
stariji.65	postotak starijih od 65 godina u ukupnom stanovništvu	2011.	DZS [42]
rac.pism	postotak stanovništva starog 10 i više godina koji su informatički pismeni (stavka korištenje Internetom)	2011.	DZS [39]
kuc.net	postotak kućanstava koja se koriste Internetom	2011.	DZS [40]
suf.def	višak prihoda, a apsolutna vrijednost ako se radi o manjku prihoda	2015.	MF [13]
tek.pror.prih.pc	tekući proračunski prihodi po stanovniku	2015.	MF [13]
rac.rash	postotak ulaganja u računalne programe i računalne usluge u ukupnim rashodima	2015.	MF [13]
dohodak.pc	prosječni dohodak po stanovniku	2010.-2012.	Regionalni razvoj[29]
izlaznost	postotak stanovnika koji su pristupili lokalnim izborima	2013.	DIP [19]
pol.konk	kategorička nominalna varijabla s 2 kategorije, poprima 1 ako se išlo u drugi krug izbora, 0 inače	2013.	DIP [19]
reizbor	kategorička nominalna varijabla s 2 kategorije, poprima 1 ako je (grado)načelnik ponovno izabran, 0 inače	2009./2013.	DIP [19], [20]

Pojašnjenje: Državni zavod za statistiku (DZS), Ministarstvo financija (MF), Državno izborno povjerenstvo (DIP)

Izvor: izrada autorice

Treća ekonomska varijabla odabrana za moguću prediktorsku su tekući proračunski prihodi po stanovniku iz 2015.

Iz 2015. su i podatci za varijablu računalni rashodi. Radi se o zbroju rashoda za računalne usluge i ulaganja u računalne programe, kao postotku ukupnih prihoda.

Podatci za varijable vezane za lokalne izore su iz 2013., kada su se izbori održali. Za vrijeme pisanja ovog rada održani su novi lokalni izbori (svibanj 2017.), no treba primijetiti da su spomenute varijable prediktorske, stoga nema smisla uzimati podatke novijeg datuma od podataka za odzivnu varijablu.

Detaljni opisi nezavisnih varijabli, razdoblje prikupljanja podataka, izvori podataka i nazivi korišteni u računalnoj obradi podataka dani su u tablici 4.2.

Ukratko, nezavisne varijable koje će se promatrati u ovom radu su socio-kulturološke (broj stanovnika, postotak visokoobrazovanih, postotak muškaraca, postotak starijeg stanovništva, postotak informatički pismenog stanovništva, postotak kućanstava s Internetom), ekonomske (odmak od proračunke ravnoteže, tekući proračunski prihodi po stanovniku, postotak ulaganja u računalne usluge i programe, dohodak po stanovniku) te političke (izlaznost na lokalne izbore, politička konkurentnost i ponovni izbor). Podatci za sve socio-kulturološke varijable prikupljeni su s internetskih stranica Državnog zavoda za statistiku i datiraju iz 2011., a podatci za političke varijable preuzeti su s internetskih stranica Državnog izbornog povjerenstva i odnose se na 2013. Podatci za dohodak po stanovniku prikupljeni su s internetskih stranica Regionalnog razvoja (2012.). Ostale ekonomske varijable jedine su aktualne - prikupljene su na internetskim stranicama Ministarstva financija s dokumenata o ostvarenju proračuna za 2015.

Poglavlje 5

Analiza i rezultati

U ovom će se poglavlju pokušati odrediti glavni prediktori razine proračunske transparentnosti lokalnih jedinica u Zadarskoj županiji. Korištene varijable detaljno su opisane u četvrtom poglavlju. Uobičajeni koraci na početku analize ove vrste su deskriptivna statistika i ispitivanje korelacije među varijablama, stoga će se ta pitanja obraditi u prva dva odjeljka ovog poglavlja. S obzirom na to da varijabla OLBI predstavlja prebrojive podatke, koristi se Poissonova regresija. U trećem odjeljku jednostrukom Poissonovom regresijom odredit će se statistička značajnost varijabli u svakom modelu posebno za određenu varijablu, a stepwise procedura će kod višestruke Poissonove regresije generirati poželjne modele u četvrtom odjeljku. U posljednja dva odjeljka uspoređuju se modeli i interpretiraju rezultati.

Za analizu se koristi besplatni softverski paket za statističku obradu podataka R (verzija 3.3.0). Sintaksa za učitavanje podataka je sljedeća:

```
podatci=read.table("varijable.txt", dec="," , header=T)
```

5.1 Deskriptivna statistika

Raspodjela vrijednosti za OLBI.2016 je dana u tablici 5.1, a deskriptivna statistika u tablici 5.2.

Deskriptivna statistika nezavisnih varijabli dana je u tablici 5.3. Jedine diskretne varijable su *pol.konk* i *reizbor*, stoga su informacije o njima predstavljene udjelima.

Tablica 5.1: Tablično prikazana raspodjela vrijednosti za OLBI.2016

Vrijednost	Frekvencija	Relativna frekvencija
0	10	0.294
1	8	0.235
2	8	0.235
3	2	0.059
4	5	0.176
5	1	0.029

Izvor: izrada autorice

Tablica 5.2: Deskriptivna statistika za varijablu OLBI.2016

Veličina uzorka	Najmanja vrijednost	Aritmetička sredina	Medijan	Najveća vrijednost	Standardna devijacija	Varijanca
34	0	1.62	1	5	1.4978	2.2433

Izvor: izrada autorice

Tablica 5.3: Deskriptivna statistika nezavisnih varijabli
(zaokruženo na 2 decimale)

Varijabla	Veličina uzorka	Aritmetička sredina	Standardna devijacija	Najmanja vrijednost	Najveća vrijednost
broj.st	34	5,015.00	12,537.49	698.00	75,062.00
visokoobraz	34	9.60	4.07	2.16	24.32
muskarci	34	50.09	1.39	44.11	52.51
stariji.65	34	20.91	4.95	13.40	35.39
rac.pism	34	48.78	6.14	36.81	65.87
kuc.net	34	37.78	7.92	25.48	57.15
suf.def	34	2,724,273.00	5,431,169.00	27,415.00	24,516,519.00
tek.pror.prih.pc	34	3,993.00	3,556.88	600.00	19,768.00
rac.rash	34	0.52	0.39	0.00	1.54
dohodak.pc	34	20,872.00	3,738.95	13,275.00	29,170.00
izlaznost	34	56.48	11.50	27.62	82.12

Varijabla	Veličina uzorka	0	1
pol.konk	34	0.59	0.41
reizbor	34	0.44	0.56

Izvor: izrada autorice

5.2 Korelacija

Uz deskriptivnu statistiku uobičajen je korak na početku analiza ove vrste ispitivanje korelacije među varijablama. Korelacija je statistički postupak za izračunavanje povezanosti dviju varijabli. Vrijednost korelacije bročano se iskazuje koeficijentom korelacije koji pokazuje u kojoj su mjeri promjene vrijednosti jedne varijable povezane s promjenama vrijednosti druge varijable.

Stoga su pronađeni koeficijenti korelacije ρ između parova nezavisnih varijabli i provedeni testovi značajnosti s nultom hipotezom $H_0 : \rho = 0$ u odnosu na alternativu $H_1 : \rho \neq 0$.

Između nekoliko uobičajenih tipova korelacije najčešće korištenih u statistici, koristi se Spearmanova korelacija jer ona ne zahtijeva nikakve pretpostavke o distribuciji podataka, varijable ne moraju biti u linearnom odnosu i nije osjetljiva na ekstremne rezultate.

Korelacijska analiza provedena je sljedećim kodom u R-u:

```
rcorr(as.matrix(podatci[ , -c(12,13,14)]), type="spearman")
```

Treba primijetiti da se analiziraju dvije binarne varijable za koje uobičajena korelacijska analiza nije prikladna, stoga se isključuju iz ovog razmatranja.

Tablica sa svim koeficijentima korelacije i njihovim značajnostima je dana u tablici 6.3 u Dodatku. Može se vidjeti da su postotak računalno pismenih i postotak kućanstva s Internetom vrlo snažno povezane varijable (0.81). U zanemarivo slabijoj vezi su visokoobrazovani i tekući proračunski prihodi po stanovniku (0.77). Umjerena povezanost je uočena između visokoobrazovanih i varijabli računalno pismenih (0.60) i kućanstava s Internetom (0.52). Nešto slabija povezanost je između te dvije varijable i starijih od 65 godina, no tu su koeficijenti razumljivo negativni (-0.50 i -0.53). Umjerena povezanost postoji i između varijabli izlaznost na izbore i broj stanovnika (-0.46), ali je opet negativnog predznaka. Posljednji par umjereno povezanih varijabli je računalna pismenost i tekući proračunski prihodi po stanovniku (0.46). Sve te povezanosti su očekivane, bilo intuitivno bilo saznanjima iz literature.

Korelacija implicira da povezane varijable mjere slične stvari. Ako su uključene u regresijski model, bit će značajne redundancije. Zadatak odabira prikladnog skupa prediktora često je zahtjevan posao, no ne očekuje se da će se varijable iz parova značajno koreliranih varijabli obje pojaviti u izabranom regresijskom modelu.

5.3 Jednostruka Poissonova regresija

Provođenjem jednostruke Poissonove regresije za svaku od nezavisnih varijabli određuje se statistička značajnost svake varijable posebno. Model za svaku nezavisnu varijablu x dan je jednostavnom relacijom

$$\log(\text{očekivanje OLBI.2016}) = \beta_0 + \beta_1 \cdot (\text{nezavisna varijabla})$$

a kod u R-u je:

```
> fit=glm(OLBI.2016~nez.var, family=poisson(link=log), data=podatci)
```

Za parametar β_1 kaže se da je statistički značajan ako se statistički značajno razlikuje od 0. Stoga su testirane sljedeće hipoteze $H_0 : \beta_1 = 0$ i $H_1 : \beta_1 \neq 0$. Za određivanje statističke značajnosti koeficijenta β_1 uzima se razina značajnosti 0.05, kao i svugdje u ovom radu.

Rezultati su prikazani u tablici 5.4. Može se uočiti da su na razini značajnosti od 0.05 statistički značajne varijable u jednostrukim modelima *broj.st*, *visokoobraz*, *suf.def*, *dohodak.pc*, *izlaznost* i *pol.konk*.

Tablica 5.4: Procjena parametara za jednostruke Poissonove modele

ime varijable	procjena parametara					
	sl.član	nagib	exp(nagib)	SE	z	p
broj.st	0.36	$1.754 \cdot 10^{-5}$	1.0000170	$6.310 \cdot 10^{-6}$	2.765	0.006
visokoobraz	-0.11	0.058	1.0598310	0.028	2.061	0.039
muskarci	3.62	-0.063	0.9391908	0.088	-0.712	0.477
stariji.65	0.46	0.001	1.0010620	0.027	0.038	0.969
rac.pism	-0.74	0.025	1.0250845	0.022	1.151	0.250
kuc.net	-0.03	0.013	1.0134989	0.017	0.789	0.430
suf.def	0.34	$4.152 \cdot 10^{-8}$	1.0000000	$1.805 \cdot 10^{-8}$	2.300	0.022
tek.pror.prih.pc	0.48	$-8.478 \cdot 10^{-8}$	0.9999999	$3.850 \cdot 10^{-5}$	-0.002	0.998
rac.rash	0.52	-0.081	0.9224977	0.359	-0.225	0.822
dohodak.pc	-1.17	$7.701 \cdot 10^{-5}$	1.0000770	$3.556 \cdot 10^{-5}$	2.166	0.030
izlaznost	2.57	-0.039	0.9619256	0.011	-3.411	0.001
as.factor(pol.konk)1	0.14	0.687	1.9875780	0.273	2.513	0.012
as.factor(reizbor)1	0.66	-0.346	0.7078040	0.270	-1.280	0.201

Izvor: izrada autorice

Provedeni su i testovi prikladnosti modela podatcima. Nulta hipoteza je da je trenutni model dobro specificiran, stoga je jedino model koji za nezavisnu varijablu uzima *izlaznost* po oba testa dobro prikladan za dane podatke. Rezultati su dani u tablici 5.5.

Tablica 5.5: GOF testovi prikladnosti modela za jednostruke Poissonove modele

ime varijable	test devijance		Pearsonov test	
	devijanica	p-vrijednost	Pearsonova statistika	p-vrijednost
broj.st	48.461	0.031	40.738	0.138
visokoobraz	50.229	0.021	41.550	0.120
muskarci	53.532	0.010	44.782	0.066
stariji.65	54.002	0.010	45.787	0.054
rac.pism	52.710	0.012	42.845	0.095
kuc.net	53.388	0.010	44.023	0.077
suf.def	49.684	0.024	42.687	0.098
tek.pror.prih.pc	54.003	0.010	45.764	0.055
rac.rash	53.952	0.010	45.792	0.054
dohodak.pc	49.392	0.025	43.022	0.092
izlaznost	42.742	0.097	33.458	0.396
as.factor(pol.konk)1	47.574	0.038	41.804	0.115
as.factor(reizbor)1	52.364	0.013	43.830	0.079

Izvor: izrada autorice

5.4 Višestruka Poissonova regresija

Višestruka analiza započinje razmatranjem modela u koji su uključene sve izabrane potencijalne prediktorske varijable. Procjena modela se dobiva sljedećim kodom u R-u:

```
> all = glm(OLBI.2016 ~ broj.st + visokoobraz + stariji.65 + muskarci
+ rac.pism + kuc.net + suf.def + tek.pror.prih.pc + rac.rash
+ dohodak.pc + izlaznost + as.factor(pol.konk)
+ as.factor(reizbor), family=poisson(link=log), data=podatci)
```

Sažetak za gornji model *all* dan je u tablici 5.6. Može se vidjeti da su u ovom modelu na razini značajnosti 0.05 značajne varijable *dohodak.pc*, *izlaznost* i *pol.konk*. Varijable su u model dodavane redom od prve navedene do zadnje.

Tablica 5.6: Sažetak za model *all*

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	6.5270	8.2392	0.79	0.4282
broj.st	-0.0000	0.0000	-1.44	0.1495
visokoobraz	-0.0251	0.1012	-0.25	0.8038
stariji.65	-0.0213	0.0605	-0.35	0.7245
muskarci	-0.1244	0.1525	-0.82	0.4149
rac.pism	-0.0705	0.0847	-0.83	0.4051
kuc.net	0.0726	0.0514	1.41	0.1581
suf.def	0.0000	0.0000	1.10	0.2713
tek.pror.prih.pc	-0.0000	0.0001	-0.16	0.8696
rac.rash	0.4996	0.5123	0.98	0.3295
dohodak.pc	0.0002	0.0001	3.16	0.0016
izlaznost	-0.0730	0.0206	-3.54	0.0004
as.factor(pol.konk)1	1.3061	0.5662	2.31	0.0211
as.factor(reizbor)1	0.4867	0.4677	1.04	0.2980

Izvor: izrada autorice

Provedeni su testovi prikladnosti modela - test devijance i Pearsonov test.

```
> pchisq(deviance(all), df.residual(all), lower.tail=F)
[1] 0.5784414
> pchisq(sum(resid(all, type="pearson")^2), df.residual(all),
  lower.tail=F)
[1] 0.3383769
```

Rezultati sugeriraju da ovaj model dobro opisuje podatke. Ipak zbog primijećene značajne korelacije među nezavisnim varijablama i rezultatima predstavljenima u odjeljku 3.2 pretpostavlja se da ovaj model nije optimalan i da će se uklanjanjem nekih varijabli dobiti bolji model.

U tablici 5.7 je prikazana analiza varijance. Ako je razlika u devijancama veća od kritične vrijednosti, tada je dodani član značajan u objašnjenju varijance u odzivu. Može se vidjeti da mnogi faktori nisu značajni.

U analizama s velikim brojem nezavisnih varijabli i inače je poželjno smanjiti broj varijabli, ali da pritom model ostane dobar za predikciju. U potrazi za boljim modelom koristit će se procedure odabira najboljeg podskupa prediktorskih varijabli objašnjene na stranici 25.

Tablica 5.7: Analiza varijance za model *all*

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			33	54.00	
broj.st	1	5.54	32	48.46	0.0186
visokoobraz	1	0.05	31	48.41	0.8148
stariji.65	1	0.31	30	48.09	0.5755
muskarci	1	0.06	29	48.03	0.8008
rac.pism	1	0.12	28	47.91	0.7284
kuc.net	1	0.27	27	47.64	0.6051
suf.def	1	0.20	26	47.45	0.6580
tek.pror.prih.pc	1	1.37	25	46.08	0.2418
rac.rash	1	0.77	24	45.30	0.3799
dohodak.pc	1	1.61	23	43.70	0.2051
izlaznost	1	15.63	22	28.07	0.0001
as.factor(pol.konk)	1	4.93	21	23.14	0.0263
as.factor(reizbor)	1	1.10	20	22.04	0.2946

Izvor: izrada autorice

Model 1

Stepwise regresija obavlja se pomoću automatizirane procedure u R-u. Može se naći u više paketa, a u ovom radu se koristi *stepAIC* iz paketa *MASS*. Unatoč imenu, *stepAIC* kao kriterij odabira modela može koristiti AIC i BIC kriterij.

Prvo se koristi kriterij BIC, a naredba je:

```
> null = glm(OLBI.2016 ~ 1, family=poisson(link=log), data=podatci)
> model1 = stepAIC(all, scope=list(lower=null, upper=all),
  direction="both", k=log(length(OLBI.2016)))
> model1.1 = stepAIC(null, scope=list(lower=null, upper=all),
  direction="both", k=log(length(OLBI.2016)))
```

Početni modeli u stepwise traženju su model *all* - model sa svim predloženim nezavisnim varijablama i model *null* - model koji sadrži samo konstantu. Skup modela koji se pretražuju određen je argumentom *scope*. To bi trebala biti ili jedna formula ili lista od dvije sastavnice - *lower* i *upper* - obje formule. Argument *direction* definira smjer stepwise traženja. *Both* određuje da se u svakom koraku dodaje po jedan prediktor ako će se kriterij poboljšati, uz mogućnost dodavanja varijable izbačene u nekom od prethodnih koraka, ili

uklanja po jedan prediktor ako će se time kriterij popraviti, uz mogućnost uklanjanja neke od prethodno dodanih varijabli.

Rezultati stepwise procedure mogu se razlikovati ovisno o modelu od kojeg se kreće. Može se vidjeti da je u oba slučaja (početni model *null* i početni model *all*) izabran model s dvije varijable: *izlaznost* i *dohodak.pc*, a koraci su dani u tablicama 5.8 i 5.9.

Tablica 5.8: Koraci stepwise procedure uz početni model *null* i kriterij BIC

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
1			33.00	54.00	120.23
2 + izlaznost	1.00	11.26	32.00	42.74	112.50
3 + dohodak.pc	1.00	8.24	31.00	34.50	107.79

Izvor: izrada autorice

Tablica 5.9: Koraci stepwise procedure uz početni model *all* i kriterij BIC

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
1			20.00	22.04	134.11
2 - tek.pror.prih.pc	1.00	0.03	21.00	22.07	130.61
3 - visokoobraz	1.00	0.11	22.00	22.18	127.20
4 - stariji.65	1.00	0.48	23.00	22.66	124.15
5 - muskarci	1.00	0.58	24.00	23.24	121.20
6 - as.factor(reizbor)	1.00	0.57	25.00	23.80	118.24
7 - rac.pism	1.00	0.51	26.00	24.31	115.23
8 - rac.rash	1.00	0.84	27.00	25.15	112.54
9 - kuc.net	1.00	2.32	28.00	27.48	111.34
10 - suf.def	1.00	3.10	29.00	30.58	110.91
11 - broj.st	1.00	1.74	30.00	32.31	109.12
12 - as.factor(pol.konk)	1.00	2.19	31.00	34.50	107.79

Izvor: izrada autorice

Sažetak odabranog modela dan je u tablici 5.10. Obje su varijable u modelu značajne na razini značajnosti od 0.05.

Tablica 5.10: Sažetak modela 1

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.9480	0.8614	1.10	0.2711
dohodak.pc	0.0001	0.0000	2.97	0.0030
izlaznost	-0.0466	0.0120	-3.89	0.0001

Izvor: izrada autorice

Testovi prikladnosti modela, dani u tablici 5.11 ukazuju na to da je model dobro prilagođen danim podatcima.

Tablica 5.11: Testovi prikladnosti modela 1

Test	Pearsonov test	Test devijance
H_0	Model je prikladan za podatke	Model je prikladan za podatke
H_1	Model nije prikladan za podatke	Model nije prikladan za podatke
p-vrijednost	0.5988	0.3039
Rezultat	Ne odbacujemo H_0	Ne odbacujemo H_0

Izvor: izrada autorice

Provodi se test omjera vjerodostojnosti za usporedbu dva modela, modela 1 i modela *all*. To je moguće jer su modeli ugniježđeni - jedan (model 1) je poseban slučaj drugoga (model *all*). Ako je razlika u devijancama veća od kritične vrijednosti, tada su dodani članovi značajni u objašnjenju varijacije u odzivu, odnosno odbacujemo nultu hipotezu da je model 1 bolji od modela *all*. Tablica 5.12 pokazuje da se na razini značajnosti 0.05 ne može odbaciti nulta hipoteza pa će se model 1 smatrati boljim od modela *all*.

Kod u R-u:

```
> anova(model1, all, test="Chisq")
```

Tablica 5.12: Test omjera vjerodostojnosti - model 1 i model *all*

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	31	34.50			
2	20	22.04	11	12.46	0.3297

Izvor: izrada autorice

Važno svojstvo Poissonove razdiobe je svojstvo jednake disperzije, stoga će se provjeriti postoje li naznake narušenosti tog svojstva. Omjer rezidualne devijance i stupnjeva slobode je blizu 1 (1.1103) pa se ne očekuje da će biti potrebe za prebacivanjem na kvazi Poissonov model. Provodi se i formalnija provjera, test preraspršenosti koji je u R-u dostupan u paketu *AER*. Rezultati ne odbacuju nultu hipotezu o jednakoj disperziji podataka.

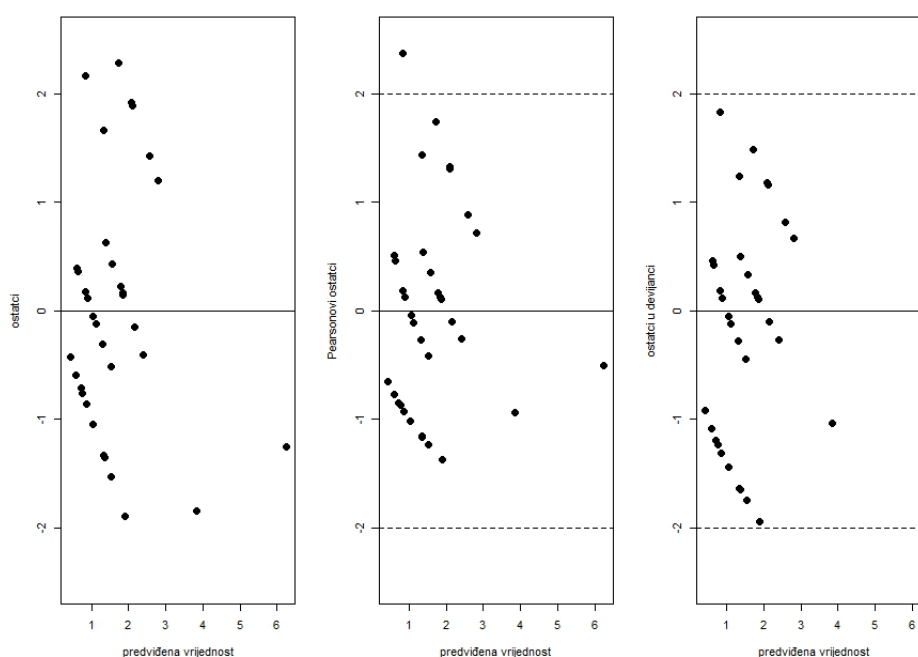
```
> dispersiontest(model1, alternative="two.sided")
```

Dispersion test

```
data: model1
z = -0.73827, p-value = 0.4604
alternative hypothesis: true dispersion is not equal to 1
sample estimates:
dispersion
0.8688442
```

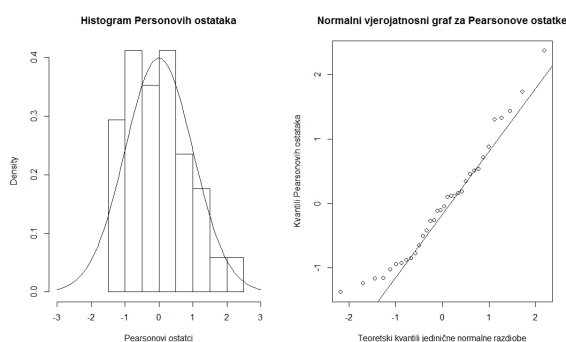
Sljedeći je korak provjera adekvatnosti modela - analiza ostataka i otkrivanje outliera i utjecajnih opažanja. Na slici 5.1 su prikazani ostaci nasuprot predviđenih vrijednosti, a na slici 5.2 je grafički ispitana normalnost histogramom i normalnim vjerojatnosnim grafom. Slika 6.1 u Dodatku ispituje postojanje nekog izrazito utjecajnog opažanja kojeg bi trebalo detaljnije proučiti.

Slika 5.1: Ostaci modela 1 nasuprot predviđenih vrijednosti



Izvor: izrada autorice

Slika 5.2: Ostatci modela 1 - grafički testovi normalnosti



Izvor: izrada autorice

Dijagnostika modela je zadovoljavajuća. Ne postoji izrazito velik ostatak, a normalni vjerojatnosni graf pokazuje tek blago odstupanje točaka od pravca. Benkovac i Zadar po utjecajnosti odskoču od drugih opažanja, ali nedovoljno da bi ih se pomnije proučavalo.

Model 2

Sada se provodi stepwise procedura krećući od modela koji sadrži samo konstantu, koristeći AIC za kriterij. Naredbe u R-u koje su predstavljene u prethodnim odjeljcima više se neće navoditi, već će se samo komentirati rezultati.

Kao najbolji je model izabran model s tri prediktora: *izlaznost*, *dohodak.pc* i *pol.konk*.

Koraci stepwise metode su dani u tablici 5.13, a sažetak odabranog modela dan je u tablici 5.14. Na razini značajnosti od 0.05 u modelu 2 značajne varijable su ponovno *izlaznost* i *dohodak.pc*.

Tablica 5.13: Koraci stepwise procedure uz početni model *null* i kriterij AIC

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
1			33.00	54.00	118.71
2 + <i>izlaznost</i>	1.00	11.26	32.00	42.74	109.44
3 + <i>dohodak.pc</i>	1.00	8.24	31.00	34.50	103.21
4 + <i>as.factor(pol.konk)</i>	1.00	2.19	30.00	32.31	103.02

Izvor: izrada autorice

Tablica 5.14: Sažetak modela 2

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.3475	0.9467	0.37	0.7136
izlaznost	-0.0383	0.0128	-2.99	0.0028
dohodak.pc	0.0001	0.0000	2.98	0.0029
as.factor(pol.konk)1	0.4430	0.2994	1.48	0.1390

Izvor: izrada autorice

Testovi prikladnosti modela u tablici 5.15 pokazuju da je model dobro prilagođen podacima.

Tablica 5.15: Testovi prikladnosti modela 2

Test	Pearsonov test	Test devijance
H_0	Model je prikladan za podatke	Model je prikladan za podatke
H_1	Model nije prikladan za podatke	Model nije prikladan za podatke
p-vrijednost	0.6494	0.3531
Rezultat	Ne odbacujemo H_0	Ne odbacujemo H_0

Izvor: izrada autorice

Tablica 5.16 prikazuje test omjera vjerodostojnosti za usporedbu modela 2 i modela 1. Može se primijetiti da se na razini značajnosti od 0.05 ne može odbaciti da je model 1 bolji od modela 2. Ipak p-vrijednost je relativno niska, stoga se model 2 neće tek tako odbaciti.

Tablica 5.16: Test omjera vjerodostojnosti - model 2 i model 1

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	31	34.50			
2	30	32.31	1	2.19	0.1388

Izvor: izrada autorice

Ispitano je postojanje naznaka narušene jednake disperzije. Omjer rezidualne devijance i stupnjeva slobode iznosi 1.0771, a tablica 5.17 pokazuje da se ne može odbaciti nulta hipoteza o jednakoj disperziji.

Tablica 5.17: Test jednake disperzije za model 2

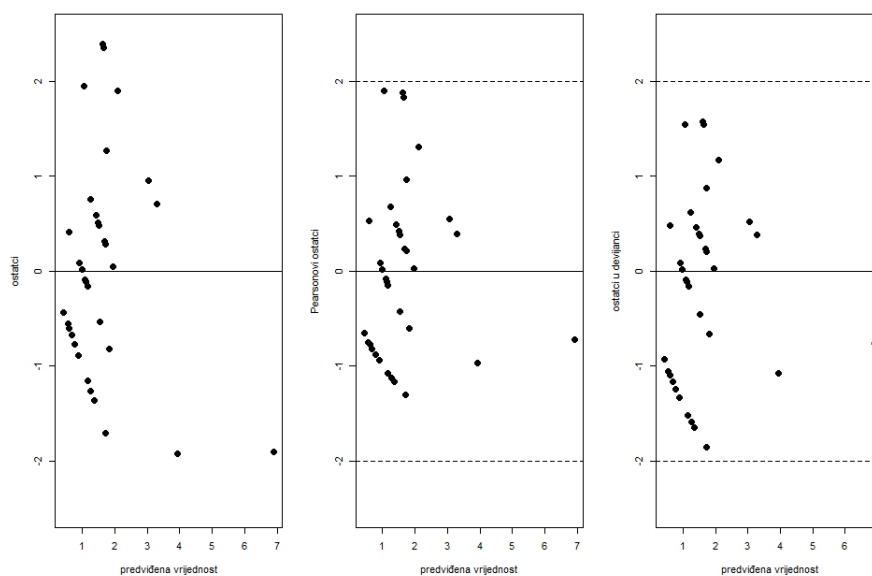
H_0	Prisutna je jednaka disperzija
H_1	Nije prisutna jednaka disperzija
p-vrijednost	0.3414
Rezultat	Ne odbacujemo H_0

Izvor: izrada autorice

Sljedeći je korak provjera adekvatnosti modela. Slika 5.3 prikazuje ostatke nasuprot predviđenih vrijednosti, a na slici 5.4 grafički je ispitana normalnost. Na slici 6.2 u Dodatku provjereno je postojanje outliera i utjecajnih opažanja.

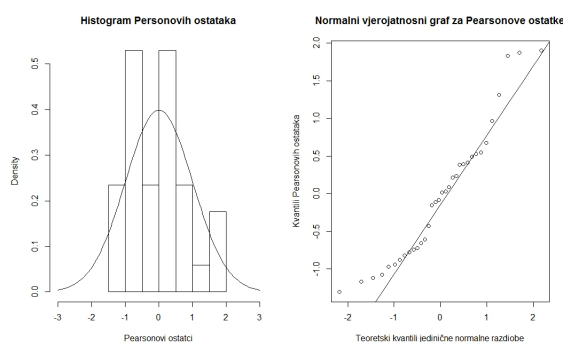
Kod histograma i normalnog vjerojatnosnog grafa mogu se uočiti određene nepravilnosti, ali rezultati su zadovoljavajući. Benkovac i Zadar ponovno se ističu po utjecajnosti, ali opet nedovoljno (vrijednosti Cookove udaljenosti su ispod 1).

Slika 5.3: Ostatci modela 2 nasuprot predviđenih vrijednosti



Izvor: izrada autorice

Slika 5.4: Ostatci modela 2 - grafički testovi normalnosti



Izvor: izrada autorice

Model 3

Stepwise procedura konačno se provodi krećući od modela *all* - modela koji sadrži sve izabrane nezavisne varijable i koristeći AIC kao kriterij.

Za najbolji se model dobiva model sa šest prediktora: *broj.st*, *kuc.net*, *suf.def*, *dohodak.pc*, *izlaznost* i *pol.konk*, a koraci procedure dani su u tablici 5.18.

Tablica 5.18: Koraci stepwise procedure uz početni model *all* i kriterij AIC

Step		Df	Deviance	Resid. Df	Resid. Dev	AIC
1				20.00	22.04	112.74
2	- tek.pror.prih.pc	1.00	0.03	21.00	22.07	110.77
3	- visokoobraz	1.00	0.11	22.00	22.18	108.88
4	- stariji.65	1.00	0.48	23.00	22.66	107.36
5	- muskarci	1.00	0.58	24.00	23.24	105.94
6	- as.factor(reizbor)	1.00	0.57	25.00	23.80	104.51
7	- rac.pism	1.00	0.51	26.00	24.31	103.02
8	- rac.rash	1.00	0.84	27.00	25.15	101.86

Izvor: izrada autorice

Sažetak odabranog modela dan je u tablici 5.19, gdje se vidi da je varijabla *kuc.net* jedina koja nije značajna na razini značajnosti 0.05.

Tablica 5.19: Sažetak modela 3

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.7791	1.0689	-0.73	0.4661
broj.st	-0.0000	0.0000	-2.68	0.0074
kuc.net	0.0409	0.0269	1.52	0.1284
suf.def	0.0000	0.0000	2.12	0.0343
dohodak.pc	0.0002	0.0000	3.32	0.0009
izlaznost	-0.0707	0.0191	-3.69	0.0002
as.factor(pol.konk)1	0.6437	0.3144	2.05	0.0406

Tablica 5.20 prikazuje testove prikladnosti modela, a rezultati su još uvjerljiviji od modela 1 i 2.

Tablica 5.20: Testovi prikladnosti modela 3

Test	Pearsonov test	Test devijance
H_0	Model je prikladan za podatke	Model je prikladan za podatke
H_1	Model nije prikladan za podatke	Model nije prikladan za podatke
p-vrijednost	0.8835	0.5658
Rezultat	Ne odbacujemo H_0	Ne odbacujemo H_0

Izvor: izrada autorice

Test omjera vjerodostojnosti koristi se za usporedbu modela 3 i modela 2, odnosno modela 3 i modela 1. U tablicama 5.21 i 5.22 mogu se uočiti relativno niske p-vrijednosti i stoga se model 3 neće odbaciti.

Tablica 5.21: Test omjera vjerodostojnosti - model 3 i model 1

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	30	32.31			
2	27	25.15	3	7.16	0.0670

Izvor: izrada autorice

Tablica 5.22: Test omjera vjerodostojnosti - model 2 i model 1

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	31	34.50			
2	27	25.15	4	9.35	0.0529

Izvor: izrada autorice

Omjer rezidualne devijance i stupnjeva slobode iznosi 0.9317, a rezultati testa preraspršenosti, koji su dani u tablici 5.23, sugeriraju odbacivanje nulte hipoteze o jednakoj disperziji. Zbog procjene parametra disperzije i vrijednosti omjera koji je ispod 1, provodi se inačica testa preraspršenosti koji testira premalu disperziju. Nulta hipoteza o preraspršenosti odbacuje se u korist alternative o premaloj disperziji, s procijenjenim parametrom disperzije od 0.6107.

Tablica 5.23: Inačice testa jednake disperzije za model 3

H_0	Prisutna je jednaka disperzija	Prisutna je jednaka disperzija
H_1	Nije prisutna jednaka disperzija	Prisutna je premala disperzija
p-vrijednost	0.0131	0.0066
Rezultat	Odbacujemo H_0	Odbacujemo H_0

Izvor: izrada autorice

Zbog premale disperzije konstruirat će se kvazi Poissonov regresijski model.

```
> model3_qp=glm(OLBI.2016 ~ broj.st + kuc.net + suf.def + dohodak.pc
+ izlaznost + as.factor(pol.konk),
family=quasipoisson(link=log), data=podatci)
```

Sažetak modela dan je u tablici 5.24. Treba primijetiti da su procijenjeni parametri isti, ali su standardne greške prilagođene za procijenjeni parametar disperzije 0.6896. Za varijablu *kuc.net* p-vrijednost se smanjila ispod 0.10.

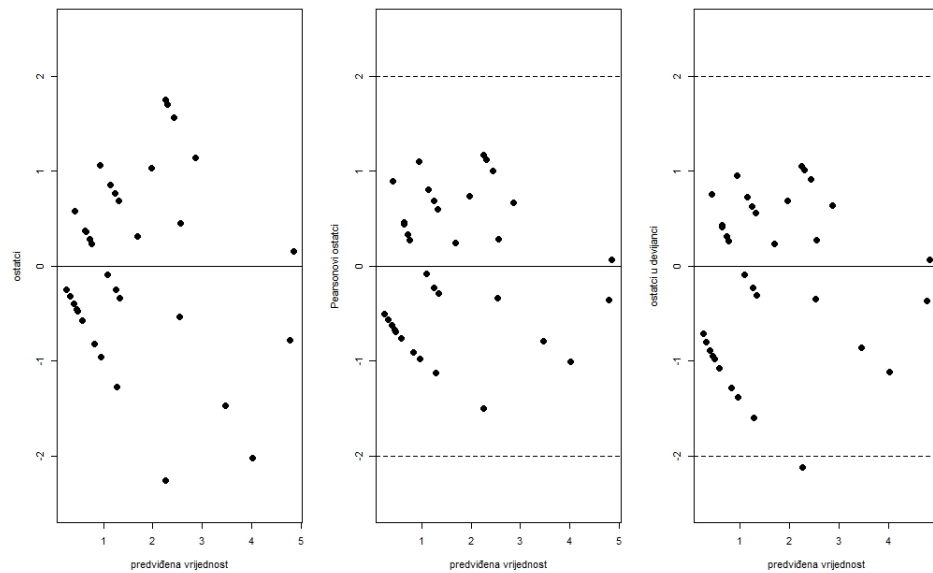
Tablica 5.24: Sažetak kvazi Poissonovog modela

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.7791	0.8876	-0.88	0.3878
broj.st	-0.0000	0.0000	-3.23	0.0033
kuc.net	0.0409	0.0224	1.83	0.0782
suf.def	0.0000	0.0000	2.55	0.0168
dohodak.pc	0.0002	0.0000	4.00	0.0004
izlaznost	-0.0707	0.0159	-4.45	0.0001
as.factor(pol.konk)1	0.6437	0.2610	2.47	0.0203

Izvor: izrada autorice

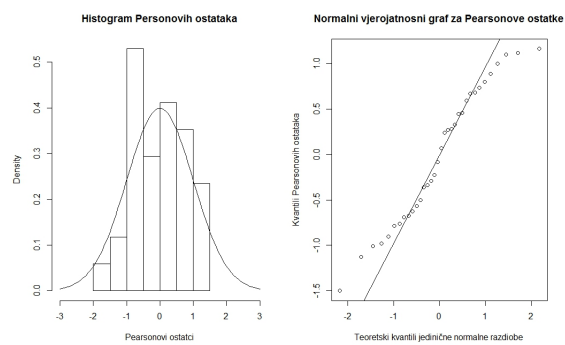
Slijedi provjera adekvatnosti modela. Na slici 5.5 prikazani su ostatci nasuprot predviđenih vrijednosti, na slici 5.6 grafički je ispitana normalnost, a slika 5.7 prikazuje detaljnu dijagnostiku utjecajnosti opažanja.

Slika 5.5: Ostatci kvazi Poissonovog modela nasuprot predviđenih vrijednosti



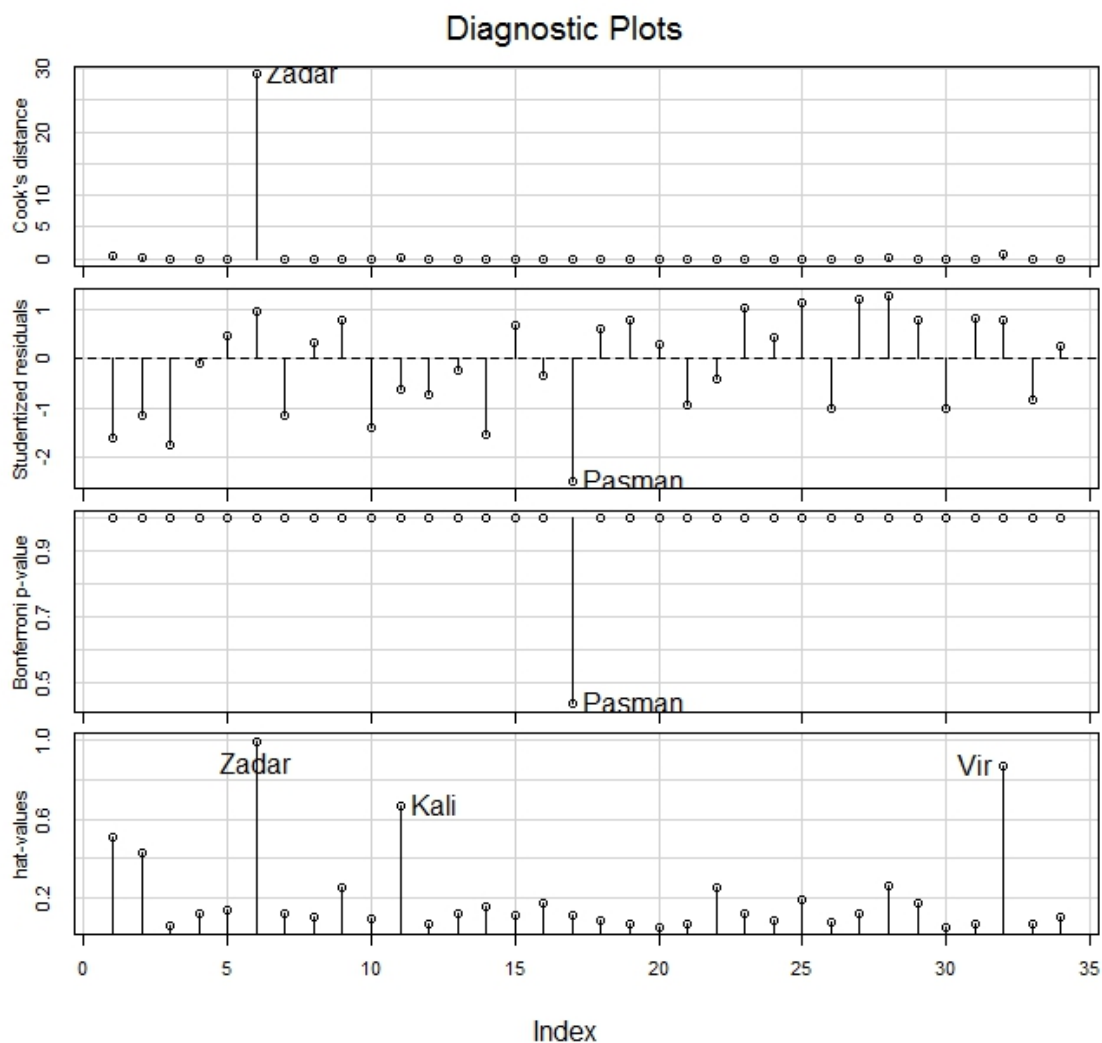
Izvor: izrada autorice

Slika 5.6: Ostatci kvazi Poissonovog modela - grafički testovi normalnosti



Izvor: izrada autorice

Slika 5.7: Dijagnostika kvazi Poissonovog modela



Izvor: izrada autorice

Ostatci se ponašaju zadovoljavajuće, ali je u ovom modelu Zadar izrazito utjecajno opažanje (vrijednost Cookove udaljenosti za Zadar iznosi gotovo 30). Grad Zadar jest posebno opažanje u uzorku - najveći je grad Zadarske županije i njezino središte. Ističe se također po sastavnicama, gradu Zadru pripadaju osim samog grada i neki udaljeniji otoci Zadarske županije (vidi [38]).

Zadar ne bi trebalo samo tako isključiti iz modela, ali ipak će se provjeriti prilagodba modela bez Zadra.

```
> m2bezzd=glm(OLBI.2016[-6] ~ broj.st[-6] + kuc.net[-6] + suf.def[-6]
+ dohodak.pc[-6] + izlaznost[-6] + as.factor(pol.konk[-6]),
family=quasipoisson(link=log), data=podatci)
```

Tablica 5.25: Sažetak kvazi Poissonovog modela bez Zadra

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0308	1.0750	-0.03	0.9774
broj.st[-6]	-0.0001	0.0001	-1.73	0.0961
kuc.net[-6]	0.0526	0.0247	2.13	0.0431
suf.def[-6]	0.0000	0.0000	2.39	0.0244
dohodak.pc[-6]	0.0001	0.0000	3.91	0.0006
izlaznost[-6]	-0.0859	0.0211	-4.06	0.0004
as.factor(pol.konk[-6])1	0.6682	0.2553	2.62	0.0146

Izvor: izrada autorice

U 5.25 se može primijetiti da su procijenjeni koeficijenti drugačiji, ali učinak varijabli ostaje istog predznaka. Za razliku od modela gdje je Zadar uključen, na razini značajnosti 0.05 u ovom modelu *broj.st* prestaje biti, a *kuc.net* postaje značajna varijabla.

5.5 Usporedba modela

Nakon prilagodbe nekoliko regresijskih modela, želi se utvrditi po čemu su ti modeli slični i u čemu se razlikuju.

Model 1:

$$\log(OLBI.2016_i) = 0.948 + 9.56 \cdot 10^{-5} \cdot dohodak.pc_i - 0.047 \cdot izlaznost_i$$

Model 2:

$$\begin{aligned} \log(OLBI.2016_i) = & 0.348 - 0.038 \cdot izlaznost_i + 9.279 \cdot 10^{-5} \cdot dohodak.pc_i \\ & + 0.443 \cdot \mathbb{1}(pol.konk_i) \end{aligned}$$

Model 3:

$$\begin{aligned}\log(OLBI.2016_i) = & -0.779 - 4.739 \cdot 10^{-5} \cdot broj.st_i + 0.041 \cdot kuc.net_i \\ & + 6.254 \cdot 10^{-8} \cdot suf.def_i + 1.542 \cdot 10^{-4} \cdot dohodak.pc_i \\ & - 0.071 \cdot izlaznost_i + 0.644 \cdot \mathbb{1}(pol.konk_i)\end{aligned}$$

Tablica 5.26: Usporedba modela

	Model 1	Model 2	Model 3
vjerodostojnost	-48.604	-47.508	-43.928
procjenjeni parametri	2	3	6
AIC	103.207	103.016	101.857
BIC	107.786	109.122	112.542
devijanica	34.504	32.314	25.155
stupnjevi slobode	31	30	27
očekivani broj 0	9.350	9.488	10.916
McFadden pseudo R^2	0.167	0.186	0.247
ML pseudo R^2	0.436	0.472	0.572
MallowCp	40.504	40.314	39.155

Izvor: izrada autorice

Tablica 5.26 pokazuje da svi kriteriji osim BIC-a najboljim modelom ocjenjuju model 3. To je očekivano jer BIC jače penalizira veći broj parametara, a model 3 uzima više nezavisnih varijabli u odnosu na modele 1 i 2.

Iako analiza varijance daje drugačije rezultate, čak preferira model 1, sa samo dva parametra, u ovom će se radu najboljim smatrati model 3. Razloga je više:

- AIC kriterij i *MallowCp* u modelu 3 su najniži, a *McFadden* i *ML pseudo R^2* najviši.
- Model 1 uzima varijable *dohodak.pc* i *izlaznost*, model 2 dodatno varijablu *pol.konk*, a one su iz 2011., što modelu smanjuje aktualnost. Model 3 je jedini koji uzima varijablu *suf.def* koju je lako pribaviti iz proračunskih dokumenata, i dostupna je za svaku godinu.
- Već na razini značajnosti od 0.10 i analiza varijance bi model 3 smatrala najboljim.

5.6 Interpretacija rezultata

Budući da model 3 ima odgovarajuće prednosti u odnosu na ostale modele, predstavljen je još jednom u tablici 5.27 kako bi se provela detaljna analiza i interpretacija rezultata.

Tablica 5.27: Sažetak modela 3 (kvazi Poisson)

	koeficijent	exp(koeficijent)	standardne greške	p-vrijednost
(Intercept)	-0.7791057	0.4588162	0.8875978	0.3878135
broj.st	-0.0000474	0.9999526	0.0000147	0.0032675
kuc.net	0.0409293	1.0417785	0.0223530	0.0781523
suf.def	0.0000001	1.0000001	0.0000000	0.0168134
dohodak.pc	0.0001542	1.0001542	0.0000386	0.0004464
izlaznost	-0.0706577	0.9317808	0.0158925	0.0001349
as.factor(pol.konk)1	0.6437133	1.9035362	0.2610495	0.0203104

Izvor: izrada autorice

Iako su neke varijable statistički značajne, nemaju veliki numerički učinak na OLBI, primjerice *broj.st*, *suf.def* i *dohodak.pc*. Za razliku od njih, varijable *izlaznost* i *pol.konk* su statistički značajne, a procijenjeni koeficijenti numerički značajni.

Tablica 5.27 pokazuje negativan učinak broja stanovnika i izlaznosti na izbore na OLBI, što je u suprotnosti s velikim brojem dosadašnjih istraživanja (vidi [16], [34], [35] za broj stanovnika, a [28] za izlaznost na izbore). Ipak, Esteller-Moré i Polo Otero su kod katalon-skih jedinica pokazali da manje lokalne jedinice imaju veću proračunsku transparentnost (vidi [11]). Među pregledanom literaturom nije pronađen rad koji pokazuje negativan utjecaj izlaznosti na izbore na OLBI, ali u velikom broju radova ta varijabla nije uopće bila značajna (vidi [35], [34]).

Za jedinično povećanje varijable *broj.st* (uz sve ostale varijable nepromijenjene) očekivano smanjenje OLBI-ja je tek 0.0047%. Koeficijent je lakše zamisliti kad kažemo da povećanje broja stanovnika za 500 osoba smanjuje OLBI za nešto više od 2 postotka (2.34%).

Utjecaj varijable *izlaznost* je izraženiji. Povećanje izlaznosti na izbore za jedan posto uz sve ostale varijable nepromijenjene smanjuje očekivani OLBI za 6.82%.

Varijabla *kuc.net* nije se pokazala statistički značajnom na razini 0.05, ali je značajna na razini značajnosti 0.10 i ima pozitivan učinak na OLBI. Za povećanje kućanstava s Internetom za jedan posto (uz sve ostale varijable nepromijenjene), očekivanje OLBI-ja poveća se za 4.18%.

Varijable *suf.def*, *dohodak.pc* i *pol.konk* imaju pozitivan učinak na OLBI, što je u skladu s postavljenim hipotezama.

Jedinično povećanje varijable *suf.def* neznatno (0.00001%) povećava očekivanje OLBI-ja, a jedinično povećanje varijable *dohodak.pc* povećava očekivanje OLBI-ja za 0.015%, naravno uz sve ostale varijable nepromijenjene. Za lakšu interpretaciju, povećanje suficita, odnosno deficita za primjerice 50000 kn, povećava OLBI za (0.5%), a povećanje dohotka po stanovniku za 100 kn povećava očekivanje OLBI-ja za 1.55%.

Što se tiče varijable *pol.konk*, očekivanje OLBI-ja je 1.90 puta veće ako je indikatorska varijabla jedinica, a ne nula.

Ukratko, analiza mogućih odrednica razine proračunske transparentnosti gradova i općina Zadarske županije djelomično je potvrdila postavljene hipoteze. Koristan uvid u odnose među nezavisnim varijablama daje korelacijska analiza, koja za neke parove varijabli pokazuje jaku povezanost. Jednostrukim Poissonovom regresijskim modelima pokazano je da su za objašnjenje zavisne varijable OLBI.2016 značajne varijable *broj.st*, *visokoobraz*, *suf.def*, *dohodak.pc*, *izlaznost* i *pol.konk*.

Višestruka analiza ističe nekoliko prilično dobrih modela, od kojih se usporedbama po više kriterija odabire "najbolji". Kod modela u koji su uključene sve nezavisne varijable, značajne su samo varijable *dohodak.pc*, *izlaznost* i *pol.konk*. Zbog primijećene korelacije među varijablama i analize varijance, model će se pokušati poboljšavati. Koristi se stepwise procedura uz različite početne modele i više kriterija odabira modela.

Prvi model sadrži dvije nezavisne varijable - *izlaznost* i *dohodak.pc* i obje su značajne. Varijabla *dohodak.pc* ima pozitivan, a *izlaznost* negativan utjecaj na OLBI.

Drugi model dodatno sadrži prediktor *pol.konk*, ali se on ne pokazuje značajnim. U oba modela provjere prikladnosti i adekvatnosti modela daju zadovoljavajuće rezultate te nema naznaka narušene jednake disperzije.

Treća opcija je model sa šest prediktora: *broj.st*, *kuc.net*, *suf.def*, *dohodak.pc*, *izlaznost* i *pol.konk*, među kojima jedino *kuc.net* nije značajna varijabla u modelu. Zbog pre-male disperzije primjenjuje se kvazi Poissonov regresijski model. Standardne su greške prilagođene za procijenjeni parametar disperzije pa se povećavaju značajnosti varijable - p-vrijednost varijable *kuc.net* spustila se ispod 0.10. Analiza prikladnosti i adekvatnosti modela uputila je na Zadar kao izrazito utjecajno opažanje, stoga se promatra i model bez grada Zadra. Procijenjeni koeficijenti su drugačiji, ali učinak varijabli ostaje istog predznaka: *broj.st* i *izlaznost* utječu negativno, dok *kuc.net*, *suf.def*, *dohodak.pc* i *pol.konk* imaju pozitivan utjecaj.

Najboljim modelom će se smatrati kvazi Poissonov model jer treći model više kriterija ocjenjuje najboljim među promatranim modelima, korištene varijable su aktualnije, a rezultati barem djelomično u skladu s dosad provedenim istraživanjima na uzorku lokalnih i područnih jedinica u RH.

Poglavlje 6

Zaključak

U ovom radu procijenjene su odrednice proračunske transparentnosti za 34 lokalne jedinice Zadarske županije. Nakon pregleda literature na temu proračunske transparentnosti, a osobito one koja se fokusira na izvještavanje putem internetskih stranica lokalne jedinice, postavljen je veći broj hipoteza temeljenih na teorijskim i empirijskim rezultatima. Potrebno je spomenuti da je to većinom strana literatura, jer su u RH istraživanja na tu temu tek u začetcima.

Promatrani indeks proračunske transparentnosti OLBI izračunao je IJF za razdoblje studeni 2015. - ožujak 2016. Indeks je kreiran tako da poprima cjelobrojne vrijednosti od 0 do 5, stoga se koristi Poissonov log-linearni model u okviru generaliziranih linearnih modela.

Nakon prilagodbe nekoliko modela, testova prikladnosti i adekvatnosti te usporedbe po više kriterija, odabran je konačan model i ti rezultati pokazuju da su odrednice razine proračunske transparentnosti gradova i općina u Zadarskoj županiji broj stanovnika, postotak kućanstava s Internetom, politička konkurentnost, izlaznost na izbore, odmak od proračunske ravnoteže i dohodak po stanovniku. Dobiveni rezultati tek djelomično podupiru postavljene hipoteze.

Prva postavljena hipoteza da veličina lokalne jedinice pozitivno utječe na OLBI iznenađujuće se nije pokazala točnom - rezultat je u suprotnosti s gotovo svim dosad provedenim studijama. Postotak kućanstava s Internetom pozitivno utječe na OLBI, kao i politička konkurentnost, što je u skladu s postavljenim hipotezama. Očekivalo se da veća izlaznost na izbore kao mjera političke aktivnosti vodi većem OLBI-ju, ali se to nije pokazalo točnim. Uzrok bi mogao biti u tome što izlazak na izbore u gradovima i općinama Zadarske županije možda nije politički motiviran, već običajno, pogotovo u manjim lokalnim jedinicama. Hipoteza da odmak od proračunske ravnoteže pozitivno utječe na OLBI je potvrđena. Stajališta o toj varijabli u teoriji su podijeljena, stoga je ovo zanimljiv rezultat unatoč malom

utjecaju. Konačno, dohodak po stanovniku pokazao se značajnom varijablom, s pozitivnim učinkom, što je u skladu sa svim pretpostavkama.

Sa stajališta metodologije, prvi nedostatak je što se radi o usko ograničenom uzorku - cijela populacija koja se promatra je Zadarska županija koju čine tek 34 lokalne jedinice, stoga se bilo kakva analiza mora vršiti na cijeloj populaciji.

Druga velika mana je vrijeme prikupljanja podataka. Većina korištenih podataka je zastarjela jer podatci godinama nisu skupljani ili objavljeni. Isto se odnosi i na OLBI koji se tek u zadnja dva ciklusa mjeri za sve lokalne jedinice u RH.

Jasno je da se može još puno napraviti u cilju usavršavanja modela i boljeg razumijevanja problematike. Buduća bi istraživanja mogla uključiti panelnu regresijsku analizu, nakon što novi podatci budu dostupni. Bilo bi korisno provesti analize odrednica proračunske transparentnosti i u drugim županijama i međusobno ih usporediti te istražiti kakav je stav i razumijevanje za proračunske dokumente među stanovnicima pojedinih gradova i općina. Trenutni OLBI mjeri tek broj objavljenih dokumenata pa bi se i sam indeks proračunske transparentnosti mogao poboljšati tako da ocjenjuje kvalitetu sadržaja, dostupnost na internetskoj stranici i tome slično.

Potrebno je još jednom naglasiti da je proračunska transparentnost lokalnih jedinica Zadarske županije vrlo niska, a posebno zabrinjava slabo objavljivanje prijedloga proračuna i objašnjenja za građane. Šteta je još veća time što je Internet široko dostupno i ekonomično sredstvo razmjene podataka. Ministarstvo financija trebalo bi bolje kontrolirati proračunske prakse i sukladno zakonima kažnjavati lokalne i područne jedinice koje ne poštuju donesene zakonske odredbe. Uz sankcije i veći pritisak javnosti, visoka bi razina proračunske transparentnosti uskoro mogla postati standard u Zadarskoj županiji, a dotad bi vlasti koje redovito objavljuju proračunske dokumente mogle profitirati zbog dobrog mišljenja koje takve prakse stvaraju kod građana.

Bibliografija

- [1] Alberto Alesina i Roberto Perotti, *Fiscal discipline and the budget process*, The American Economic Review **86** (1996), br. 2, 401–407.
- [2] James E Alt, David Dreyer Lassen i Shanna Rose, *The causes of fiscal transparency: evidence from the US states*, IMF Staff papers (2006), 30–57.
- [3] Bojan Basrak, *Materijali s predavanja iz Financijskog praktikuma*, https://web.math.pmf.unizg.hr/~bbasrak/?Courses_%2F_Nastava:Financijski_praktikum1, pristupljeno: 30.06.2017.
- [4] Mihaela Bronić, *Kodeks fiskalne transparentnosti*, Povremeno glasilo IJF (2011), br. 53.
- [5] José Caamaño-Alegre, Santiago Lago-Peñas, Francisco Reyes-Santias i Aurora Santiago-Boubeta, *Budget transparency in local governments: an empirical analysis*, Local Government Studies **39** (2013), br. 2, 182–207.
- [6] Carmen Caba Pérez, Manuel Pedro Rodríguez Bolívar i Antonio M López Hernández, *e-Government process and incentives for online public financial information*, Online Information Review **32** (2008), br. 3, 379–400.
- [7] A Colin Cameron i Pravin K Trivedi, *Regression analysis of count data*, sv. 53, Cambridge university press, 2013.
- [8] Samprit Chatterjee i Jeffrey S Simonoff, *Handbook of regression analysis*, sv. 5, John Wiley & Sons, 2013.
- [9] Roger Debreceeny, Glen L Gray i Asheq Rahman, *The determinants of Internet financial reporting*, Journal of accounting and public policy **21** (2002), br. 4-5, 371–394.
- [10] Annette J Dobson i Adrian Barnett, *An introduction to generalized linear models*, CRC press, 2008.

- [11] Alejandro Esteller-Moré i Jose Polo Otero, *Fiscal Transparency: (Why) does your local government respond?*, Public Management Review **14** (2012), br. 8, 1153–1173.
- [12] A Esteller-Moré i J Polo Otero, *Analysis of the causes of fiscal transparency: Evidence from Catalanian municipalities*, unpublished paper (2008).
- [13] Ministarstvo financija, *Ostvarenje proračuna JLP(R)S za period 2014. - 2016.*, <http://www.mfin.hr/hr/ostvarenje-proracuna-jlprs-za-period-2014-2016>, pristupljeno: 16.08.2017.
- [14] International Monetary Fund, *Fiscal transparency*, <http://www.imf.org/external/np/fad/trans>, pristupljeno: 28.06.2017.
- [15] Juan L Gandía i Maria C Archidona, *Determinants of web site information by Spanish city councils*, Online Information Review **32** (2008), br. 1, 35–57.
- [16] Maria Dolores Guillamón, Francisco Bastida i Bernardino Benito, *The determinants of local government's financial transparency*, Local Government Studies **37** (2011), br. 4, 391–406.
- [17] Cheng Hsiao, *Analysis of panel data*, br. 54, Cambridge university press, 2014.
- [18] IJF, *Interaktivna karta razine transparentnosti proračuna*, <http://www.ijf.hr/transparentnost/gradovi/>, pristupljeno: 10.08.2017.
- [19] Državno izborno povjerenstvo, *Informacija o izborima članova predstavničkih tijela jedinica lokalne i područne (regionalne) samouprave i općinskih načelnika, gradonačelnika i župana te njihovih zamjenika, 2013.*, http://www.izbori.hr/izbori/dip_ws.nsf/public/index?open&id=BF0A&, pristupljeno: 16.08.2017.
- [20] Državno izborno povjerenstvo, *Informacija o izborima članova predstavničkih tijela jedinica lokalne i područne (regionalne) samouprave i općinskih načelnika, gradonačelnika i župana te njihovih zamjenika, 2009.*, <http://www.izbori.hr/izbori/izbori09.nsf/wi?openform>, pristupljeno: 16.08.2017.
- [21] M Kent Jennings, *Gender roles and inequalities in political participation: Results from an eight-nation study*, Western Political Quarterly **36** (1983), br. 3, 364–385.
- [22] Fawzi Laswad, Richard Fisher i Peter Oyelere, *Determinants of voluntary Internet financial reporting by local government authorities*, Journal of Accounting and Public Policy **24** (2005), br. 2, 101–121.
- [23] Russ Lavery, *An Animated Guide: An Introduction To Poisson Regression*, <http://www.lexjansen.com/nesug/nesug10/sa/sa04.pdf>, pristupljeno: 10.08.2017.

- [24] Liang Ma i Jiannan Wu, *What drives fiscal transparency? Evidence from provincial governments in China*, (2011).
- [25] Peter McCullagh i James A Nelder, *Generalized Linear Models*, no. 37 in *Monograph on Statistics and Applied Probability*, 1989.
- [26] Međunarodni monetarni fond, *Code of Good Practices on Fiscal Transparency*, <https://www.imf.org/en/Publications/Policy-Papers/Issues/2016/12/31/Code-of-Good-Practices-on-Fiscal-Transparency-revised-April-2007-PP4175>, pristupljeno: 10.08.2017.
- [27] K Ott, M Bronić, M Petrušić i B Stanić, *Proračunska transparentnost županija, gradova i općina: studeni 2015. – ožujak 2016.*, *Povremeno glasilo IJF* (2016), br. 107.
- [28] Suzanne J Piotrowski i Gregg G Van Ryzin, *Citizen attitudes toward transparency in local government*, *The American Review of Public Administration* **37** (2007), br. 3, 306–323.
- [29] Ministarstvo regionalnog razvoja, *Vrijednosti indeksa razvijenosti i pokazatelja za izračun indeksa razvijenosti 2013.*, <https://razvoj.gov.hr/o-ministarstvu/djelokrug-1939/regionalni-razvoj/indeks-razvijenosti/vrijednosti-indeksa-razvijenosti-i-pokazatelja-za-izracun-indeksa-razvijenosti-2013/3214>, pristupljeno: 16.08.2017.
- [30] Hrvatski sabor, *Zakon o izmjenama i dopunama Zakona o proračunu*, http://narodne-novine.nn.hr/clanci/sluzbeni/2015_02_15_277.html, pristupljeno: 28.06.2017.
- [31] Hrvatski sabor, *Zakon o proračunu*, http://narodne-novine.nn.hr/clanci/sluzbeni/2003_06_96_1216.html, pristupljeno: 16.08.2017.
- [32] Hrvatski sabor, *Uredba o indeksu razvijenosti*, http://narodne-novine.nn.hr/clanci/sluzbeni/2010_05_63_1999.html, pristupljeno: 10.08.2017.
- [33] Hrvatski sabor, *Zakon o pravu na pristup informacijama*, http://narodne-novine.nn.hr/clanci/sluzbeni/2013_02_25_403.html, pristupljeno: 16.08.2017.
- [34] Carlos Serrano-Cinca, Mar Rueda-Tomás i Pilar Portillo-Tarragona, *Factors influencing e-disclosure in local public administrations*, *Environment and planning C: Government and Policy* **27** (2009), br. 2, 355–378.

- [35] Daniel Albalade del Sol, *The institutional, economic and social determinants of local government transparency*, Journal of Economic Policy Reform **16** (2013), br. 1, 90–107.
- [36] Alan K Styles i Mack Tennyson, *The accessibility of financial reporting of US municipalities on the Internet*, Journal of Public Budgeting, Accounting & Financial Management **19** (2007), br. 1, 56.
- [37] Ministarstvo uprave, *Lokalna i područna (regionalna) samouprava*, <https://uprava.gov.hr/o-ministarstvu/ustrojstvo/uprava-za-politicki-sustav-drzavnu-upravu-te-lokalnu-i-podrucnu-regionalnu-samoupravu/lokalna-i-podrucna-regionalna-samouprava/842>, pristupljeno: 28.06.2017.
- [38] Odbor za Statut i Poslovnik, *Statut grada Zadra*, <http://krsevan.grad-zadar.hr/GradskoVijece/VA%C5%BDNI%20DOKUMENTI/STATUT/Statut%20Grada%20Zadra%20%20-%20pro%C4%8Di%C5%A1%C4%87eni%20tekst.pdf>, pristupljeno: 13.08.2017.
- [39] Državni zavod za statistiku, *Stanovništvo staro 10 i više godina prema informatičkoj pismenosti, starosti i spolu, popis 2011.*, https://www.dzs.hr/Hrv/censuses/census2011/results/htm/h01_01_34/h01_01_34_zup13_5207.html, pristupljeno: 16.08.2017.
- [40] Državni zavod za statistiku, *Privatna kućanstva prema posjedovanju osobnog računala i korištenju Interneta, popis 2011.*, https://www.dzs.hr/Hrv/censuses/census2011/results/htm/h02_01_20/H02_01_20.html, pristupljeno: 16.08.2017.
- [41] Državni zavod za statistiku, *Stanovništvo staro 15 i više godina prema najvišoj završenoj školi, starosti i spolu, popis 2011.*, https://www.dzs.hr/Hrv/censuses/census2011/results/htm/h01_01_32/h01_01_32.html, pristupljeno: 16.08.2017.
- [42] Državni zavod za statistiku, *Kontingenti stanovništva po gradovima/općinama, popis 2011.*, https://www.dzs.hr/Hrv/censuses/census2011/results/htm/H01_01_03/H01_01_03.html, pristupljeno: 16.08.2017.

Sažetak

Cilj ovog rada je naći odrednice proračunske transparentnosti gradova i općina u Zadarskoj županiji. Promatrani indeks proračunske transparentnosti OLBI mjeri broj objavljenih ključnih proračunskih dokumenata na internetskoj stranici lokalne jedinice.

Hipoteze postavljene u skladu s postojećom literaturom su empirijski testirane na uzorku sve 34 lokalne jedinice u Zadarskoj županiji - 6 gradova i 28 općina.

Korišten je Poissonov log-linearni model unutar okvira generaliziranih linearnih modela jer je varijabla od interesa prebrojiva.

Hipoteze su djelomično potvrđene. Rezultati ukazuju da postotak kućanstava s Internetom, politička konkurentnost, odmak od proračunske ravnoteže i dohodak po stanovniku zaista imaju pozitivan utjecaj na OLBI. Za veličinu lokalne jedinice i izlaznost na izbore dobiveni su iznenađujući rezultati - u Zadarskoj županiji utjecaj tih varijabli je negativan. Ostale varijable se nisu pokazale značajnima u objašnjenju razine OLBI-ja.

Vrijednost ovom radu daje to što je prvo u RH koje analizira određenu županiju, a ne lokalne jedinice na razini cijele države. Također, tema proračunske transparentnosti još nije dovoljno zastupljena u javnosti pa bi pisanje o tome moglo potaknuti javnost da se više zanima za proračunske dokumente, znanstvenike da proučavaju različite aspekte ovog problema te vladajuće da više poštuju osnovno pravo svakog građanina da dobije uvid u što se troši njegov novac kao poreznog obveznika.

Summary

The main goal of this thesis is to find determinants of budget transparency of cities and municipalities in Zadar County. Observed budget transparency index OLBI counts the number of released key budget documents on the local unit's website.

Hypotheses, set in accordance with existing literature, are empirically tested on sample of 34 local units of Zadar County - 6 cities and 28 municipalities.

Poisson log-linear model is used within the frame of generalized linear models because the variable of interest is countable.

Hypotheses are partially confirmed. Results indicate that percentage of households with Internet access, political competitiveness, departure from budget balance and income per capita indeed have a positive impact on OLBI. Surprising results were obtained for size of the local unit and voter turnout - in Zadar County those variables' effect is negative. Other variables have not been proved significant in explanation of OLBI level.

This thesis is valuable because it is the first in the Republic of Croatia that analyzes specific county, and not local units at the level of the whole country. Also, the topic of budget transparency is not yet sufficiently represented in public so writing about it may encourage: the public to be interested in budget documents, the scientists to study different aspects of this problem and the authorities to be more respectful towards the basic right of every citizen as a taxpayer to get an insight into where his money is spent.

Dodatak

Tablica 6.1: Karakterizacija najčešćih razdioba u eksponencijalnoj familiji

	Normalna	Poisson	Binomna	Gama
Oznaka	$N(\mu, \sigma^2)$	$P(\mu)$	$B(m, \pi)/m$	$G(\mu, \nu)$
Raspon od y	$(-\infty, \infty)$	$0(1)\infty$	$\frac{0(1)m}{m}$	$(0, \infty)$
Parametar disperzije: ϕ	$\phi = \sigma^2$	1	$1/m$	$\phi = \nu^{-1}$
Kumulativna funkcija: $b(\theta)$	$\theta^2/2$	e^θ	$\log(1 + \exp \theta)$	$-\log(-\theta)$
$c(y; \phi)$	$-\frac{1}{2}(\frac{y^2}{\phi} + \log(2\pi\phi))$	$-\log y!$	$\log \binom{m}{my}$	$\nu \log(\nu y) - \log y - \log \Gamma(\nu)$
$\mu(\theta) = \mathbb{E}(Y; \theta)$	θ	$\exp(\theta)$	$e^\theta/(1 + e^\theta)$	$-1/\theta$
Kanonska veza: $\theta(\mu)$	identiteta	log	logit	recipročna vrijednost
Funkcija varijance: $V(\mu)$	1	μ	$\mu(1 - \mu)$	μ^2

Izvor: autorica prema [25]

Tablica 6.2: Oblici devijance za najčešće razdiobe iz eksponencijalne familije, uz sumaciju $i = 1, \dots, n$

Razdioba	Devijanca
Normalna	$\sum (y - \hat{\mu})^2$
Poissonova	$2 \sum \{y \log(y/\hat{\mu}) - (y - \hat{\mu})\}$
Binomna	$2 \sum \{y \log(y/\hat{\mu}) + (m - y) \log[(m - y)/(m - \hat{\mu})]\}$
Gama	$2 \sum \{-\log(y/\hat{\mu}) + (y - \hat{\mu})/\hat{\mu}\}$

Izvor: autorica prema [25]

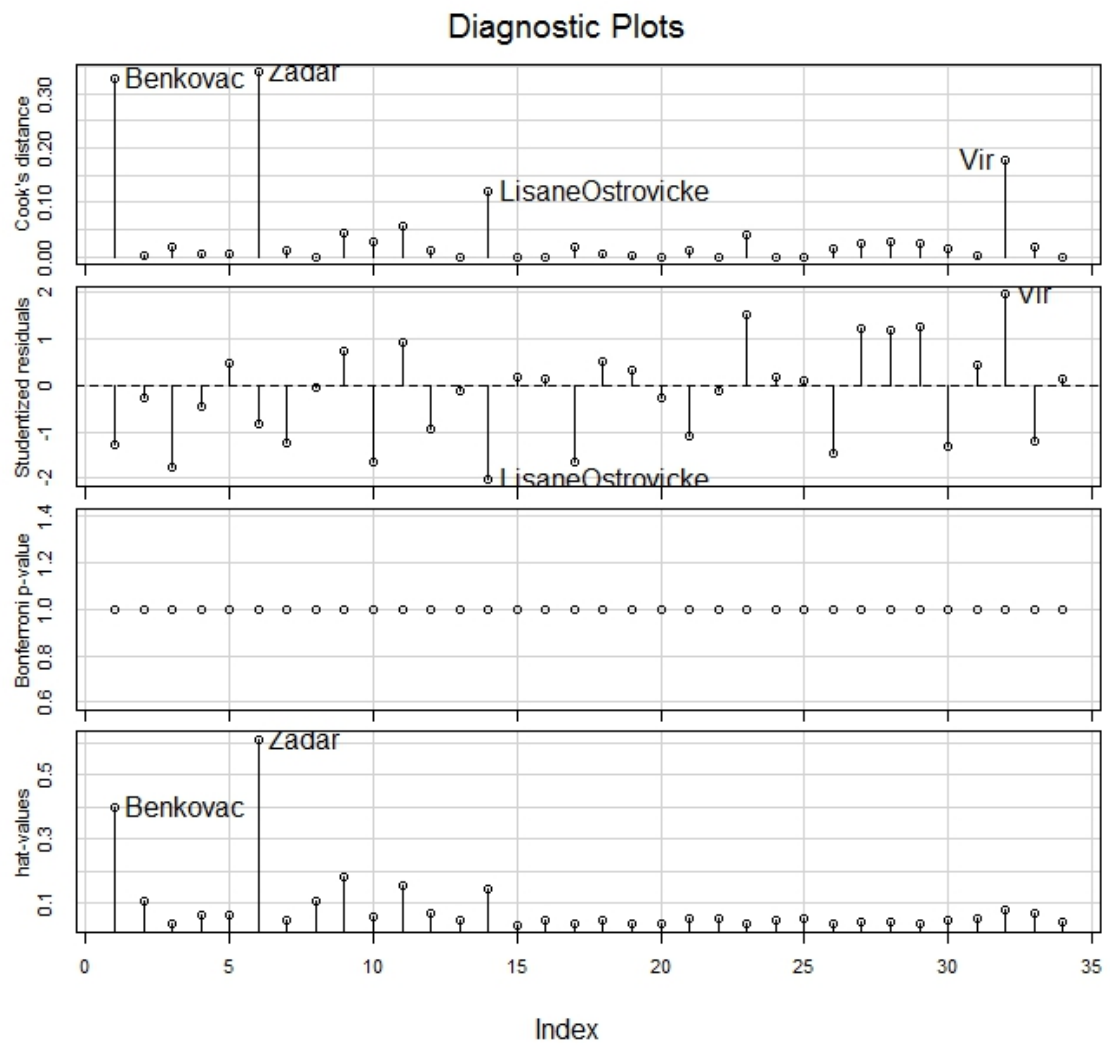
Tablica 6.3: Spermanovi koeficijenti korelacije i njihove značajnosti

	suf.def	tek.pror.prih.pc	dohodak.pc	rac.rash	broj.st	stariji.65	muskarci	visokoobraz	rac.pism	kuc.net
suf.def	0.25									
tek.pror.prih.pc	0.06	0.16								
dohodak.pc	0.11	0.33	0.05							
rac.rash	0.27	-0.16	-0.07	-0.10						
broj.st	0.02	0.20	-0.03	0.28	-0.34					
stariji.65	0.09	-0.15	-0.19	-0.05	-0.23	0.13				
muskarci	0.17	0.77 * **	0.30	0.10	0.16	0.11	-0.20			
visokoobraz	0.00	0.46 * *	0.21	0.06	0.22	-0.50 * *	-0.26	0.60 * **		
rac.pism	0.06	0.37 *	0.30	0.10	0.17	-0.53 * *	-0.35 *	0.52 * *	0.81 * **	
kuc.net	-0.26	0.22	0.27	0.11	-0.46 * *	-0.17	0.04	0.19	0.33	0.40 *

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

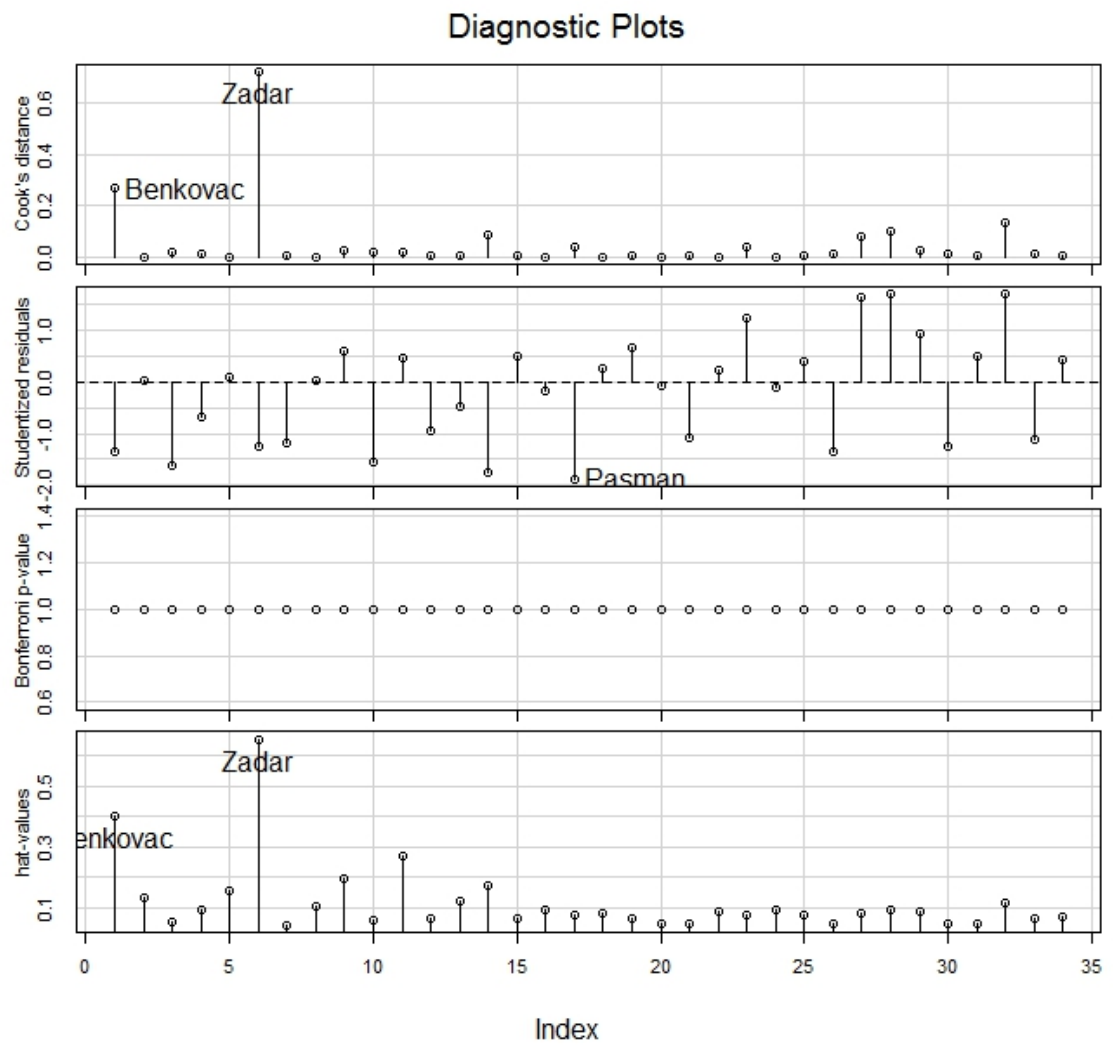
Izvor: izrada autorice

Slika 6.1: Dijagnostika modela 1



Izvor: izrada autorice

Slika 6.2: Dijagnostika modela 2



Izvor: izrada autorice

Životopis

Marija Višić rođena je 21. kolovoza 1992. u Zadru, a odrasta na otoku Ugljanu. Nakon Osnovne škole Valentin Klarin koju pohađa u Ugljanu, a potom u Preku, daljnje školovanje nastavlja u Gimnaziji Jurja Barakovića u Zadru, opći smjer. Akademske godine 2011./2012. upisuje preddiplomski sveučilišni studij Matematike na Prirodoslovno-matematičkom fakultetu u Zagrebu, te nakon njegovog završetka akademske godine 2013./2014. upisuje diplomski sveučilišni studij Financijska i poslovna matematika na istom fakultetu.